論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Hand-Eye Calibration and Terrain Mapping under Extreme Light Conditions using Stereo Vision Camera for Humanitarian Demining Robot
著者(和文)	LiJianhua
Author(English)	Jianhua Li
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第9972号, 授与年月日:2015年9月25日, 学位の種別:課程博士, 審査員:遠藤　玄,鈴森　康一,大熊　政明,小田　光茂,塚越　秀行,福島 E　文彦
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第9972号, Conferred date:2015/9/25, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Hand-Eye Calibration and Terrain Mapping under Extreme Light Conditions using Stereo Vision Camera for Humanitarian Demining Robot

A thesis presented

by

Jianhua Li

to

The Department of Mechanical and Aerospace Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Engineering

Academic Advisors:

Associate Professor Gen Endo

Professor Edwardo F. Fukushima

Tokyo Institute of Technology

Tokyo, Japan

September 2015

Thesis supervisors
Associate Professor Gen Endo
Professor Edwardo F. Fukushima

Author
Jianhua Li

Hand-Eye Calibration and Terrain Mapping under Extreme Light Conditions using Stereo Vision Camera for Humanitarian Demining Robot

Abstract

Stereo vision camera is widely used in the robotics area for terrain mapping, object detection, object classification, navigation, self-localization and so on. This thesis presents the problems that a Humanitarian Demining Robot Gryphon meets in the field when a stereo vision camera is used for terrain mapping: accurate Kinematic calibration and Hand-Eye calibration and ensuring the 3D terrain model still could be well acquired with a stereo vision camera even in the extreme light conditions. These problems are also the common and important issues when applying stereo vision camera in field and this thesis focuses on them. This thesis introduces a new calibration method that performs simultaneous kinematic calibration and hand-eye calibration based on a traditional method that uses a sequence of pure rotations of the manipulator links. The new method considers an additional joint angle constraint, which improves the calibration accuracy especially when the circular arc that can be measured by the stereo vision camera is very limited. In addition, instead of using existing lighting enhancement methods such as Exposure Fusion to increase the texture of 2D images, this thesis describes a new stereo matching method that is directly done using the images grabbed with multiple exposures. The 2D image process of exposure

fusion is not needed. Experimental results using a manipulator developed for humanitarian demining demonstrate that with the proposed method the relative errors between the end-effector and the external points mapped by the stereo vision camera are greatly reduced. Through the experiments in laboratory and outdoors with a stereo vision camera fixed on a tripod and held in the hand, it is verified that the proposed method consistently allowed more valid points to be obtained and the 3D terrain model could be built more accurately. Especially when the local window-based method was used, the proposed method performed much better than the traditional methods. Field experiments are planned to be conducted with the Gryphon system in Angola in the near future to further evaluate the proposed methods. The methods proposed and presented in this thesis can be used in other robot platforms.

Acknowledgments

I would like to thank Emeritus Professor Shigeo Hirose, Professor Edwardo F. Fukushima and Professor Gen Endo for providing me the chance of studying in Hirose-Fukushima Laboratory, which currently is Suzumori-Endo Laboratory. I am happy to be in this laboratory and spend precious time with the laboratory members. Having the opportunity to study in this laboratory is a great advancement in my personal knowledge.

Thanks to Professor Edwardo F. Fukushima for giving me lots of useful advice and great help in my research. I gradually learn how to do research, how to find problems and how to solve problems. Without his support and guidance, I would not be able to accomplish this research.

Thanks to Professor Gen Endo for giving me great help in my research. Especially for my preparing for my graduation, Professor Gen Endo supports me a lot and gives me important and useful suggestion.

Thanks to Professor Koichi Suzumori for giving me great help and suggestion in my research.

Thanks to Emeritus Professor Shigeo Hirose for giving me lots of useful suggestion in my research.

Thanks to all the laboratory members for your kindly help to my research and my living in Japan. Thanks to Dr. Alex Masuo Kaneko, Mr. Lin Lin, Dr. Yu Chun Fu, Dr. Arturo Ceron, Dr. Hiroaki Ishida and all the other members for your kindly help to me. Thanks to Dr. Allan Ben for correcting my English writing.

Great thanks to the Ministry of Education, Culture, Sports, Science and Technology (MEXT) for providing the financial support for my study program. It is my great honour to get the MEXT Scholarship.

Finally, my deepest gratitude to my father, my mother and my sister for your constant support to me. My dear wife's support and help is important for me. Great thanks to my family members.

Contents

List of Figures

List of Tables

List of Abbreviations

ATV	All Terrain Vehicle
UXO	Unexploded Ordnance
MMD	Mine Metal Detector
2D	Two Dimensional
3D	Three Dimensional
FOV	Field of View
HFOV	Horizontal Field of View
LMA	Levenberg-Marquardt Algorithm
RANSAC	Random Sample Consensus
DOF	Degrees of Freedom
RMS	Root Mean Square
SURF	Speeded Up Robust Features
HDR	High Dynamic Range
SGM	Semi-Global Matching
IMU	Inertial Measurement Unit
GPS	Global Positioning System

List of Key Words

Kinematic calibration: The goal is to determine the spatial kinematic relationship between the joints and links and improve the accuracy of the kinematic model of the manipulator.

Hand-Eye calibration: The goal is to find the position and orientation of the sensor relative to the manipulator.

Extreme Light Conditions: From outdoor shade to outdoor sunlight, the scene luminance could be changed from 100 lux to 100000 lux.

Chapter 1

Introduction

Stereo vision camera is widely used in the robotics area for terrain mapping, object detection, object classification, navigation, self-localization and so on. This thesis presents the problems that a Humanitarian Demining Robot Gryphon meets in the field when a stereo vision camera is used for terrain mapping: accurate Kinematic calibration and Hand-Eye calibration and ensuring the 3D terrain model still could be well acquired with a stereo vision camera even in the extreme light conditions. These problems are also the common and important issues when applying stereo vision camera in field and this thesis focuses on them. The methods proposed and presented in this thesis can be used in other robot platforms.

1.1 Description of the Humanitarian Demining Robot Gryphon

Humanitarian demining is the action of clearing a land area of mines and unexploded ordnance (UXO) to allow local population to safely return living there. In manual demining operations, a human deminer systematically scans the ground with a mine metal detector (MMD). This process is time consuming, expensive and can be dangerous to the deminers. The actual clearing of a minefield is a very risky task even for highly trained professionals. Most automatic demining systems tend to explode the ordnances

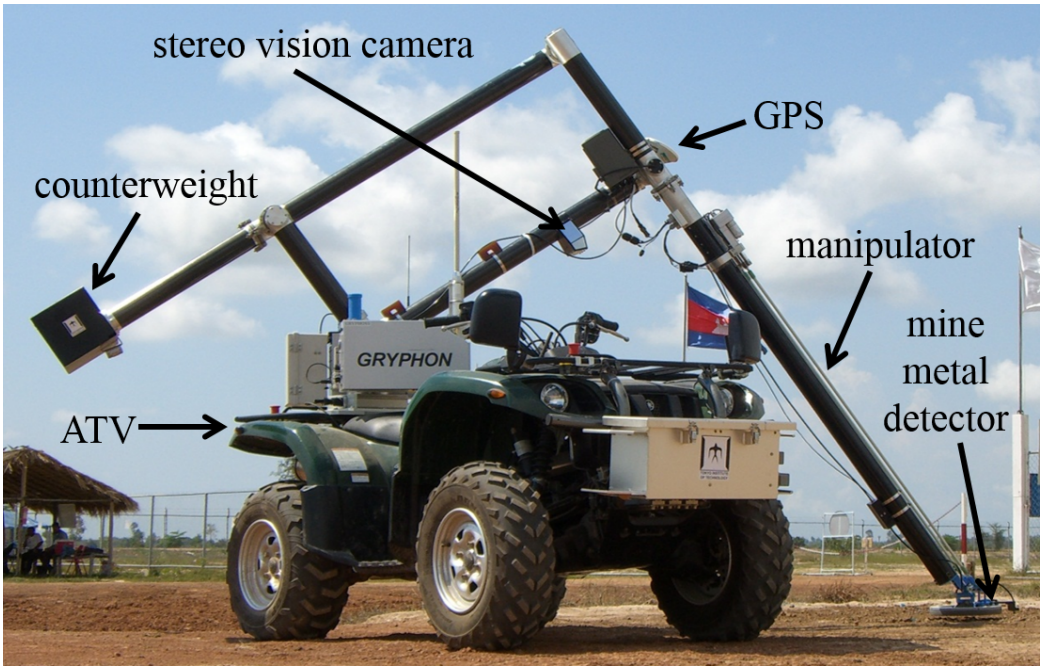


Figure 1.1: Overview of Gryphon platform.

without defusing them and this method is not totally safe and always be followed by a careful manual inspection afterwards. A mechanical system to assist human deminers, presented in Figure 1.1 and named Gryphon [1], has been developed at the Tokyo Institute of Technology since 2002. Gryphon platform is in fact an advanced mine detector, which is able to autonomously scan the interested area (namely a lane), and mark the suspected spots with a marking system or internally for a later inspection.

The terrains in which it has to operate are very inhomogeneous in their characteristics, as they can be hard, muddy, sandy, and with temperatures ranging from below zero levels to very hot and humid weather. The control system of the Gryphon platform was developed to be as easy as possible to be used, because it is not feasible to use highly trained engineers in the field. This leads to the fact that Gryphon platform was intended to be used by personnel with a minimum basic training, working for governments of humanitarian agencies. The consequence for the fact is that the entire platform must be cheap, robust, and easy to maintain. The robot also ensures

Figure 1.2: Remote Control Unit.

safe operations because the operator never needs entering the minefield, as neither the robot. In fact, Gryphon platform is able to perform scanning from the outside of the minefield, thanks to its long arm.

As presented in Figure 1.1, the Gryphon platform consists of an all-terrain vehicle (ATV), mounted with a robotic manipulator that carries a MMD. It can be remotely supervised by an operator with a Remote Control Unit as shown in Figure 1.2. To ensure high detection rate, the MMD must be scanned as close as possible to the soil, so a precise calibration of the system is needed for ensuring precise positioning. It is expected to achieve a positioning error smaller than 20.0 mm.

Laser range finder and stereo vision camera have been widely used for terrain mapping and obstacle avoidance in robotics area for many years. In [2], a comparison of the advantages and limitations of stereo vision camera

Figure 1.3: Point Grey $Bumblebee^{TM}$ stereo vision camera.

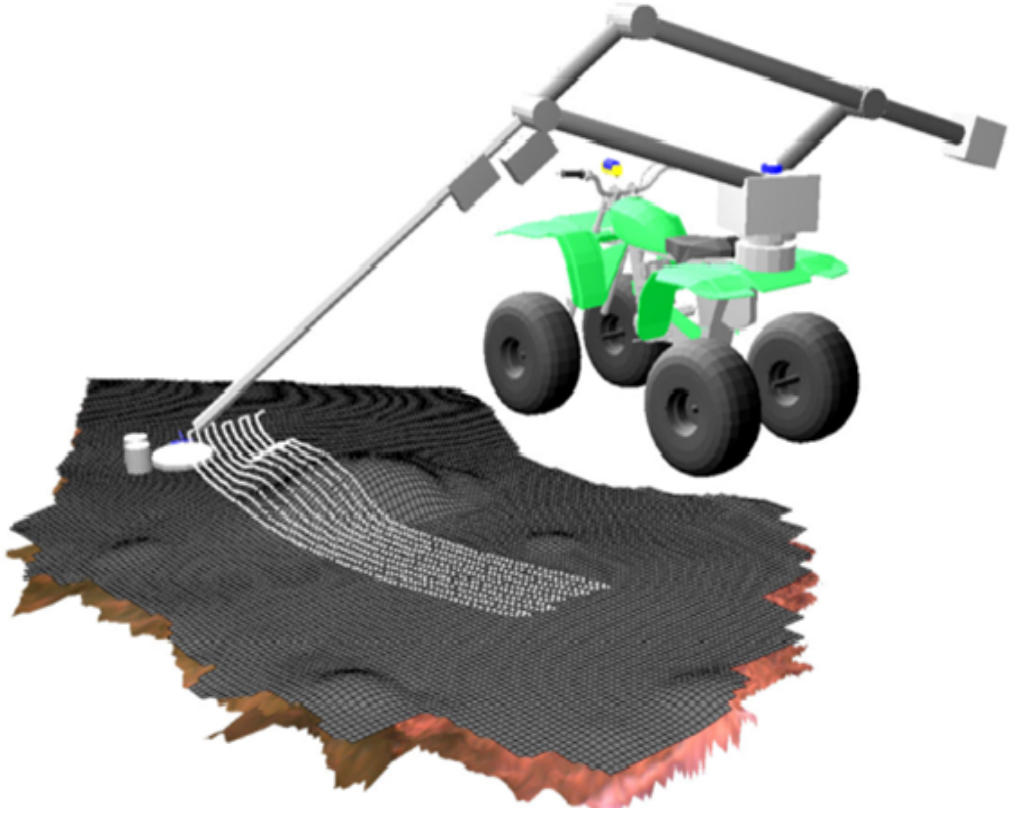


Figure 1.4: The automatically scanning operation based on the height map.

and laser range finder was presented.

First, the characters of laser range finder are listed [2].

- Its accuracy decreases only slightly with range and it is typically more accurate than stereo vision camera. The user calibration is not needed.

- It is insensitive to lighting and can operate at night or in sunlight. However, it can be affected by direct sunlight.

- It requires little processing power and outputs range values directly.

- Its scan is produced incrementally, so it is susceptible to errors during rapid motion. Besides, since it contains sensitive optics and mirrors, it is susceptible to shock and vibration problems.

- It produces only a 2D cross section and no color information is returned.

- Its range measurements are sparse compared to stereo vision camera. Besides, its measures may give 0 meters (bad or no reflection) or be ambiguous (like lower parts of moving chairs) [3].

- It is effective when the environment has unique physical structure or shape. Unfortunately, the laser range finder with limited range can fail around homogeneous building structures such as long corridors [4].

- It is active sensing and requires power. Laser beams could be detected by others.

Next, the characters of stereo vision camera are presented [2].

- Its range accuracy decreases with range and it has limited field of view. The user calibration is required.

- It depends on adequate lighting and texture of the scene, and requires the environment to contain unique visual features based on variable appearance.

- It requires significant processing for correspondence matching and is computationally intensive to work with.

- It produces an "instantaneous" snapshot of the environment and does not have the moving parts. So it is more robust to shock and vibration.

- It produces dense range information and is particularly useful for providing rich 3D information. Color information can be associated with range data.
- It is passive sensing and no energy is emitted into the environment.

Considering robustness to various field conditions, cost, precision and processing speed, an industrial stereo vision camera, the Point Grey $Bumblebee^{TM}$ stereo vision camera, was selected. This stereo vision camera is shown in Figure 1.3, and it is based on a rigid aluminum body to avoid shocks and vibrations to move the two inner cameras from their perfect parallel alignment. As presented in Figure 1.1, the stereo vision camera is attached to the manipulator link and used to map the terrain. As shown in Figure 1.4, the 3D terrain model can be built using the stereo vision camera and the Humanitarian Demining Robot Gryphon can automatically scan the interested area with a MMD at a constant distances from the ground.

1.2 Problem description

The Gryphon platform shown in Figure 1.1 is proven to be better than human operators. However, there are still some issues needed to be solved and the improvements are necessary.

1.2.1 Background and motivation of the Kinematic calibration and Hand-Eye calibration for the Humanitarian Demining Robot

For the Gryphon platform, a precise kinematic calibration and hand-eye calibration is fundamental to scan as close as possible to the soil and it is expected to achieve a positioning error smaller than 20.0 mm. An initial calibration may be performed in a laboratory/factory environment using customized calibration tools and/or high-precision external sensors, but the implementation of an in-field self-calibration method is still necessary for the following main reasons.

1. System health-status auto-check: to assure high reliability of the system throughout its lifetime use, a self-calibration method can be used to compare the joint initial angles and camera position and orientation parameters calibrated from self-calibration and from the original laboratory/factory pre-calibration.

2. Re-calibration after in-field maintenance: re-calibration is needed when the counterweight holder fixing position is changed, or the stereo vision camera is replaced.

   (a) Initial position errors: the manipulator joint angles are measured by incremental encoders that are initialized with the manipulator's counterweight at its resting position, which is defined to be the counterweight holder fixing position, and small initial positioning errors cause large errors in the manipulator's tip.

   (b) Counterweight holder assembly precision: the counterweight holder fixing position can slightly differ from one ATV to another.

   (c) Stereo vision camera position and orientation errors: the stereo vision camera can be replaced by other stereo vision camera, or its position and orientation can be different from one manipulator to another. Small stereo vision camera position and orientation errors cause large errors in terrain mapping.

In fact, the "re-calibration after in-field maintenance"is a major challenge if it is assured high precision positioning without using a high-precision external sensor.

1.2.2 Terrain mapping in extreme light conditions

Since no assumption of the ground's shape can be made a priori, it is necessary that a system is able to perform 3D acquisition of the scene even in extreme light conditions. The luminance of an object is measured in lux or candelas per square meter ($1lux = 1cd/m^2$). The range of the luminance of an object is called the dynamic range and it is defined as the ratio of

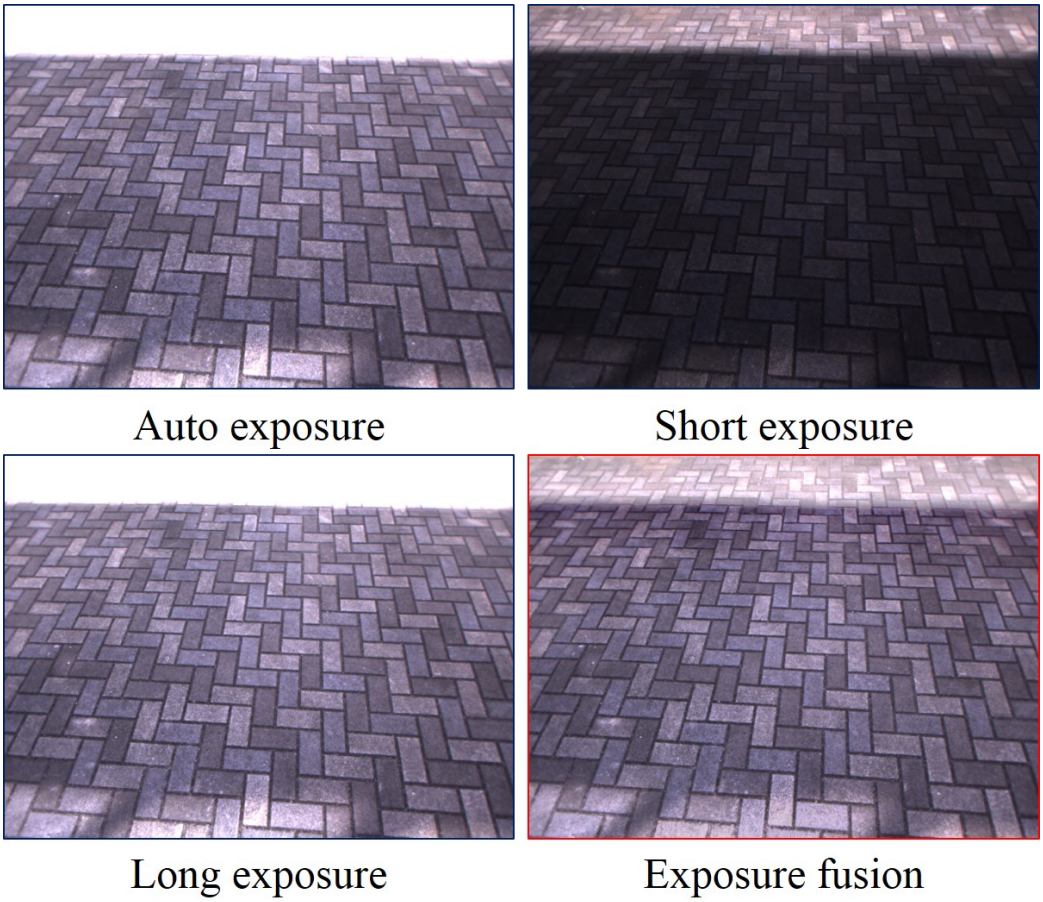


Figure 1.5: The scene with extreme light conditions.

the maximum luminance value to the minimum luminance value within the specimen. Intensity values of scenes in the real world can have a very broad dynamic range, and from outdoor shade to outdoor sunlight, the scene luminance could be changed from 100 lux to 100000 lux. For a digital camera it is possible that the image acquired with auto exposure saturates in some areas while keeping others visibly underexposed. This is particularly true for scenes that have areas of both low and high illumination. As shown in the auto exposure image of Figure 1.5, the upper part is overexposed. The limitation comes from the camera's dynamic range, which represents the limits of luminance range that a given device can capture. The dynamic range of cameras is limited by the charge-coupled devices (CCD), analog-to-digital

conversion (ADC) and film characteristics [32]. For some lighting conditions in field, the stereo correspondence algorithm is unable to find enough features to perform a depth analysis. The scan can become totally impossible or even more, the manipulator is wrongly positioned and hits obstacles or mines. The problem was reproduced in laboratory and reported in [33]. The depth map calculated from the stereo pair lacked so many features that only a small fraction of it was actually used to compute the 3D information.

1.3 Related work

The related work is surveyed and briefly described in this section.

1.3.1 Kinematic calibration and Hand-Eye calibration

The purpose of robot calibration is to improve robot accuracy through software rather than changing the mechanical structure or design of the robot and minimize the risk of having to change application programs due to slight changes or drifts (such as wear of parts, dimensional drifts and tolerances, and component replacement effects) in the robot system [5]. The robot calibration is a deeply treated subject with an extensive list of supporting publications and a good overview of robot calibration was given in [5]. In an effort to classify most of the approaches for the robot calibration, three levels of robot calibration are defined.

Level 1 Calibration. Its goal is to ensure that the reading from a joint sensor yields the correct joint displacement.

Level 2 Calibration. Its goal is to improve the accuracy of the kinematic model of the manipulator as well as the relationship between the joint transducers and the actual joint displacement.

Level 3 Calibration. it is defined as "non-kinematic" calibration. At this level, a number of effects exist which may be modelled. For example, deflection of the robot links or the effects of joint compliance and gear backlash may be significant.

In the robot calibration, the internal reference point fixed on the robot

is used and its position could be obtained from sensor. This thesis mainly focuses on the **Level 2 Calibration**, which is also referred to as "Kinematic calibration" and whose objective is to determine the spatial kinematic relationship between the joints and links [5]. Determining these parameters can be undertaken with two different procedures. The first corresponds to open or closed loop methods, which are based on minimizing the errors between the measured positions by sensor and the predefined positions of the robot. The second consists of the screw axis measurement methods, which seek to accurately determine the real positions of the robot's articulation axes, to later obtain the kinematic parameters by means of the algebraic relations existing between the axes [6]. Many robot calibration methods have already been studied in the literature, such as: nonlinear least squares estimation [7], plane constraint [8], interpolation of joint angles considering joint compliance [9], consideration of non-geometric parameters [10], pure rotation of the links and circle-point analysis technique [7, 12, 13].

In [7], a nonlinear least squares estimation with Levenberg-Marquardt Algorithm was introduced, where the parameters are obtained through minimizing the errors between the measured positions and the predicted positions based on the kinematic model. Neural network has also been used by several researches for robot calibration [9], [10]. In [9], it was used for interpolating the relationship between joint angles and their errors due to joint compliance. In [10], a laser tracking system was employed for measuring robot arm's tip with high accuracy and residual errors caused by non-geometric parameters were further reduced by using neural network. In [11], the authors presented an automatic approach for the kinematic calibration of the humanoid robot NAO.

In [12–15], pure rotation method, or circle-point analysis method, was used for Robot calibration. This method was introduced in detail in [7]. For the circle-point analysis method, it involves freezing all joints except one, and rotating that joint through a range of angles. The resulting end-effector positions nominally describe an arc, whose center point lies on the rotational axis of that joint and whose normal vector describes the direction of the rotational axis [13]. A collection of such arcs uniquely defines the parameters

of the robot. In [13], a high precision SMX laser-tracking coordinate measuring machine was used to successfully calibrate an industrial robot and it was shown that through the measurements obtained using the circle-point analysis technique and the highly accurate SMX coordinate measuring device an industrial robot was successfully calibrated. In [16], with the circle-point analysis method, the authors obtained a combination of real kinematic and dynamic parameters which describe the robots movement, improving its precision with a physical understanding of the errors.

Whenever a sensor such as a stereo vision camera is mounted on a robot manipulator, it is important to know the relationship between the sensor and manipulator. The problem of determining this relationship is referred to as the hand-eye calibration problem [17]. The main task of hand-eye calibration is to find the position and orientation of the sensor relative to the manipulator [18]. A vast amount of literature is available on the topic of hand-eye calibration. In the hand-eye calibration, the external reference point in the calibration frame is used and its position could be obtained from sensor. There are two main approaches to estimate the hand-eye transformation. One is the classical formulation $AX = XB$, which is based on "move the hand and observe/perceive the movement of the eye"and another approach is to use $AX = YB$, which is based on "simultaneous estimation of the hand-eye transformation and the pose of the robot in the world" [18]. In [19], $AX = XB$ was solved by separating the problem into its orientational component and positional component. A linear solution was made in [20] to solve $AX = YB$, where rotation quaternion was used for the rotation part. In [21], manifolds such as circles were used for the hand-eye calibration using a relatively large number of poses to estimate a single circle, improving the overall accuracy of calibration. In [22], the authors presented several formulations of Hand-Eye calibration that lead to multivariate polynomial optimization problems, and it was shown that the method of convex linear matrix inequality relaxations can be used to effectively solve these problems and obtain globally optimal solutions.

Manufacturing/assembly precision of mechanical parts, initial positioning and/or sensor reading errors can cause undesirable end-effector tip position-

ing errors in robotic manipulators. This is also the case for the Gryphon platform. With a sequence of pure rotations of the manipulator links, the real positions of the robot's articulation axes could be accurately determined, and the Kinematic calibration and Hand-Eye calibration can be performed simultaneously. However, in field it is not practical to calibrate the Gryphon platform with a high-precision external sensor, and the self-calibration with the stereo vision camera which is already attached to the manipulator link of the Gryphon platform is necessary. Former methods cannot perform as well as expected when the joint motion that can be measured is limited by the Field of View (FOV) of camera.

1.3.2 Stereo matching under extreme light conditions

By using a reduced exposure time, one may sacrifice lowlight detail in exchange for improved detail in areas of high illumination and this is demonstrated in the short exposure image of Figure 1.5. Similarly, by increasing exposure time, a better representation of lowlight areas may be gotten, at the cost of losing information in areas of high illumination and an example of this is shown in the long exposure image of Figure 1.5. Through multiple exposures, the dynamic range of images could be increased. The 2D images captured under multiple exposures are processed with two known techniques: HDR imaging [35] and Exposure Fusion [36]. In [35], the algorithm uses these differently exposed photographs to recover the response function of the imaging process, up to factor of scale, using the assumption of reciprocity. With the known response function, the algorithm can fuse the multiple photographs into a single, high dynamic range radiance map whose pixel values are proportional to the true radiance values in the scene. In [36], the authors proposed a technique for fusing a bracketed exposure sequence into a high quality image, without converting to HDR first. This avoids camera response curve calibration and is computationally efficient. In [37], an approach was presented to improve the effective dynamic range of cameras by using multiple photographs of the same scene taken with different exposure times. In [38], a image-based technique was proposed for enhancing the

shape and surface details of an object and the input to this system is a small set of photographs taken from a fixed viewpoint, but under varying lighting conditions.

In the real application, it is possible that the camera is moved when the images are grabbed with multiple exposures and it becomes crucial to properly align the input images before fusing a high dynamic range (HDR) image. Image registration is the process of overlaying images (two or more) of the same scene taken at different times, from different viewpoints, and/or by different sensors. The registration geometrically align two images (the reference and sensed images) [40]. A review of the classic image registration methods is presented in [40]. There are two major image alignment algorithms: pixel-based methods and feature-based alignment methods. The image alignment and image stitching algorithms are reviewed in [41]. Feature-based approaches have the advantage of being more robust against scene movement, and are potentially faster if implemented the right way [41]. In [43], the image alignment was formulated as a multi-image matching problem and the invariant local features were used to find matches between all of the images. Since the invariant local features are used to detect the key-points, this method is insensitive to the ordering, orientation, scale and illumination of the input images and can recognise multiple panoramas in an unordered image dataset. In [45], the direct featureless projective parameter estimation approach was applied to image resolution enhancement and compositing, and the pairs of images were registered. The simultaneous registration of multiple images was proposed in [46]. In [44], a method for alignment of images acquired by sensors of different modalities was presented. In [47], a computationally simple method was proposed to jointly estimate the registration parameters and the parameters describing the exposure correction, directly from the image intensity values. In [48], after capturing high dynamic range images from a set of photographs taken at different exposures, the key-points or feature-points in these images were searched. The key-points were used to find matrices, which transform a set of images to a single coordinate system.

Determination of three dimensional (3D) data from images is of central importance in the field of machine vision and one of the most direct way

of achieving this from image data is stereo vision. Stereo matching is one of the most active research areas in computer vision [49]. In [50], a taxonomy of dense, two-frame stereo methods was presented and this taxonomy is designed to assess the different components and design decisions made in individual stereo algorithms. The stereo algorithms generally perform (subsets of) the following four steps [50]:

- Matching cost computation;
- Cost (support) aggregation;
- Disparity computation or optimization;
- Disparity refinement.

Normally, there are two different stereo matching algorithms: local window-based method and global method. For the local window-based algorithms, the disparity computation at a given point depends only on intensity values within a finite window. In [51], the authors propose an area-based local stereo matching algorithm for accurate disparity estimation across all image regions. In [52], a real-time local stereo matching using guided image filtering was presented and it is reported that the GPU implementation of the stereo algorithm can process stereo images with a resolution of 640 x 480 pixels and a disparity range of 26 pixels at 25 fps. The global algorithms, which make explicit smoothness assumptions and then solve an optimization problem, typically seek a disparity assignment that minimizes a global cost function that combines data and smoothness terms [50]. The global stereo matching methods, such as Graph Cuts [54] and Simulated Annealing [55], have been proposed. In [53], a GPU-based stereo matching system was presented.

One of the biggest problems of applying stereo vision techniques in field robotics is how to acquire 3D terrain maps under extreme light conditions. Through multiple exposures, the dynamic range of images could be increased. A system architecture was introduced in [34] for terrain mapping using stereo vision camera. Traditionally, the 2D images captured with multiple exposures are processed with exposure fusion [36]. With the resulting fused images, the disparity image is calculated with stereo matching method and the 3D terrain

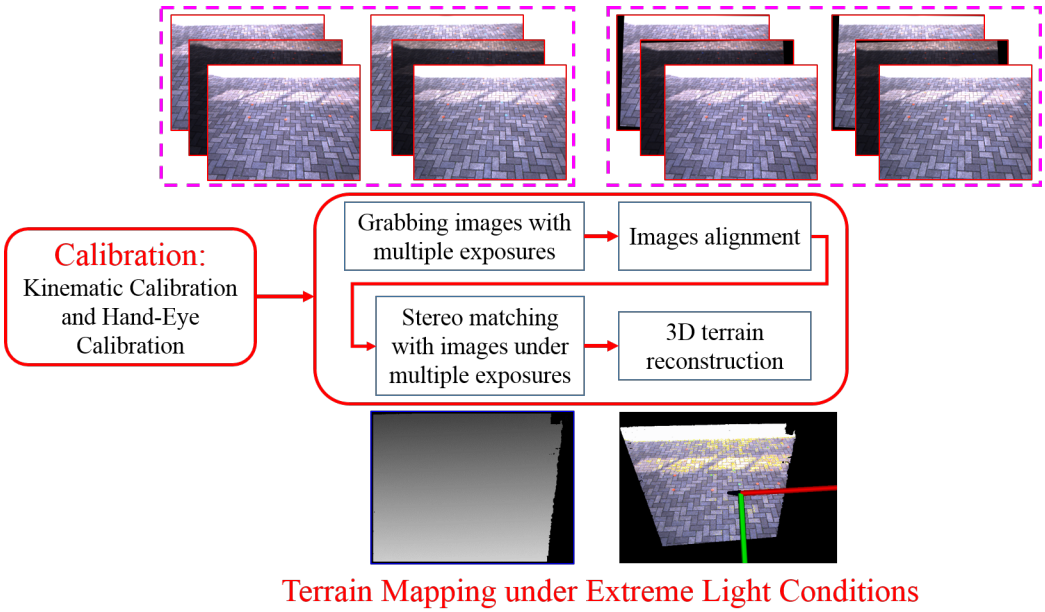


Figure 1.6: The main research work of the thesis.

is reconstructed. However, it is possible that some information is lost when the images are fused with exposure fusion and the stereo matching result with the fused images is not as good as expected.

1.4 Thesis outline and contribution

The main research work of this thesis is presented in Figure 1.6. This thesis mainly focuses on the method of Kinematic calibration and Hand-Eye calibration and the terrain mapping under extreme light conditions. The method introduced in this thesis uses a sequence of pure rotations of the manipulator links and the Kinematic calibration and Hand-Eye calibration are performed simultaneously. Previous methods do not have a good performance when the joint motion that can be measured is limited by the FOV of camera, and/or no external sensor is available, which is the case studied in this thesis. In this thesis, the joint angle, which can be measured by a high-precision sensor, is added as a constraint to fit a circular arc and compared to the previous methods the calibration accuracy is improved. Through multiple

exposures, the dynamic range of images could be increased. Instead of using existing lighting enhancement methods such as Exposure Fusion to increase the texture of 2D image, in this thesis the stereo matching was directly done using the images grabbed with multiple exposures and the 2D image process of exposure fusion is not needed. Compared to the previous methods, the 3D terrain model could be built more accurately with the proposed method.

First, Chapter 2 introduces the proposed Kinematic calibration and Hand-Eye calibration method: Pure Rotation through Fitting Circular Arc in 2D Space with Joint Angle Constraint. In this Chapter, a self-calibration scheme using the stereo vision system already included in the system was introduced, and the Kinematic calibration and Hand-Eye calibration were performed simultaneously. The introduced calibration methodology relies on a sequence of pure rotations of the manipulator links, while tracking the manipulator's tip and an external arbitrary fixed reference point by the stereo vision camera. The new method considers an additional joint angle constraint, which improves the calibration accuracy especially when the circular arc that can be measured by the stereo vision camera is very limited. Experimental results using a manipulator developed for humanitarian demining demonstrate that with the proposed method the relative errors between the end effector and the external points mapped by the stereo vision camera are greatly reduced.

Next, Chapter 3 presents the proposed Hand-Eye calibration method: Pure Rotation through Directly Fitting Circular Arc in 3D Space with Joint Angle Constraint. A camera mounted on a Pan-Tilt unit requires the so called Hand-Eye calibration to calibrate its assembling position and orientation. In the literature, the method that uses a sequence of pure rotations through the pan motion and tilt motion was presented to solve the Hand-Eye calibration problem. However, the calibration accuracy degrades when the circular arc that can be measured by the camera is quite limited. This Chapter extends the circular arc fitting method described in Chapter 2 that solves this problem by adding a joint angle constraint, and introduces a new method that directly fits a circular arc in 3D space with joint angle constraint. Compared to the traditional methods, simulations results showed the improved performance of the proposed methods, and experimental results

using a Pan-Tilt-Camera system confirmed that with the proposed methods a better calibration result can be obtained.

Then, Chapter 4 introduces the algorithm of Terrain Mapping under Extreme Light Conditions with Direct Stereo Matching Method. Through multiple exposures, the dynamic range of images could be increased. Since it is possible that the camera is moved when the images are grabbed with multiple exposures, the images grabbed with short and long exposures were aligned to the image captured with auto exposure. In this Chapter, instead of using existing lighting enhancement methods such as exposure fusion to increase the texture of 2D image, the stereo matching is directly done using the images captured with multiple exposures and the matching costs of the images grabbed with multiple exposures are directly summed by weight. Compared with the previous methods such as exposure fusion, it is not necessary to fuse the images grabbed with multiple exposures, and for each pixel of the matching image the local information in its local window acquired from the images grabbed with multiple exposures could be better retained. In order to evaluate the performance of the proposed method, two different stereo matching algorithms were used: a local window-based method and semi-global method. Through the experiments in laboratory and outdoors with a stereo vision camera fixed on a tripod and held in the hand, it was verified that the proposed method consistently allowed more valid points to be obtained and the 3D terrain model could be built more accurately. Especially when the local window-based method was used, compared to the traditional methods, the proposed method performed much better.

Finally, Chapter 5 presents the conclusion of this thesis. A summary of the achievements and the future work are shown. Field experiments are planned to be conducted with the Gryphon system in Angola in the near future to further evaluate the proposed methods. The proposed methods can be used in other robot platforms.

Chapter 2

Kinematic Calibration and Hand-Eye Calibration with Pure Rotation Method by Fitting Circular Arc in 2D Space with Joint Angle Constraint

The pure rotation calibration method relies on a sequence of rotations of the manipulator links, while using a camera to track an internal fixed reference point on the manipulator's tip and an external arbitrary reference point fixed in the robot coordinate frame. In the camera coordinate frame, the resulting reference point positions measured by camera are nominally located in a plane whose normal vector is the direction of the rotational axis and describe a circular arc whose center lies on the rotational axis of that joint. The process to identify the calibration parameters of the kinematic calibration and hand-eye calibration is divided into two procedures.

- *Procedure 1.* Using the resulting reference point positions measured by camera in each single-joint rotation, the normal vector, center and

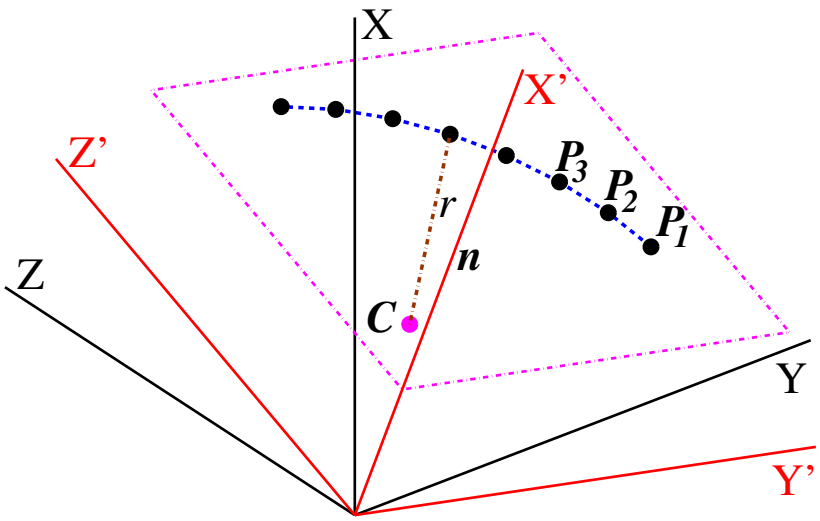


Figure 2.1: Best-fit plane estimation in the 3D space. In the camera coordinate frame, For the circular arc, $\boldsymbol{n}$ is its estimated normal vector, $\boldsymbol{C}$ is its estimated center and r is its estimated radius.

radius of the circular arc are estimated in the camera coordinate frame.

- *Procedure 2.* Based on the estimated normal vectors, centers and radii of all these single-joint rotations, and using their geometry relationship to the calibration parameters, the calibration parameters are solved [7, 13, 21].

For the pure rotation calibration method, it is critical to accurately estimate the normal vector, center and radius of the circular arc. In this Chapter, the focus is mainly on the *Procedure 1.*

2.1 Fitting a circular arc in 2D space with joint angle constraint

As shown in Figure 2.1, in 3D space, given $N(N \geq 3)$ points $\boldsymbol{P}_t$, $t = 1, 2, \ldots, N$, which are nominally located in a plane that can be defined by a normal vector $\boldsymbol{n}$ and a signed distance d from the origin, a circular arc is

constructed and defined by three parameters: normal vector $\boldsymbol{n}$, center $\boldsymbol{C}$ and radius r.

Note: $\boldsymbol{P}_t = [P_{xt} \quad P_{yt} \quad P_{zt}]^T$, the Euclidean norm $|\boldsymbol{P}| = \sqrt{{P_x}^2 + {P_y}^2 + {P_z}^2}$, $\boldsymbol{C} = [C_x \quad C_y \quad C_z]^T$, $\boldsymbol{n} = [n_x \quad n_y \quad n_z]^T$ with $\boldsymbol{n}^T\boldsymbol{n} = 1$.

Traditionally it is divided into two steps to fit a circular arc in 3D space [7, 13].

- *Step 1.* Fitting the resulting reference point positions to an optimal 3D plane to compute the normal vector $\boldsymbol{n}$ of the circular arc in a least-squares sense.
- *Step 2.* Projecting the resulting reference point positions onto the optimal plane and within the identified 2D plane fitting the projected point positions to a circular arc to estimate the center $\boldsymbol{C}$ and radius r of the circular arc in a least-squares sense.

2.1.1 Best-fit plane estimation

In each single-joint rotation, as shown in Figure 2.1, point $\boldsymbol{P}_t$ is nominally located in a plane which can be defined by a normal vector $\boldsymbol{n}$ and a signed distance d from the origin. With Random Sample Consensus (RANSAC) [25], the best-fit plane is estimated and the normal vector $\boldsymbol{n}$ is calculated. When fitting the plane with RANSAC, the optimization problem defined with (2.1) and solved with Least Squares Method is used for the parameters estimation and the point-plane distance error DE_t defined with (2.2) is used for error estimation.

$$\min \sum_{t=1}^{N} \left(\boldsymbol{n}^T\boldsymbol{P}_t + d\right)^2 \tag{2.1}$$

$$DE_t = |\boldsymbol{n}^T\boldsymbol{P}_t + d| \tag{2.2}$$

The distance threshold DT is defined with (2.3). S_d is the scale factor of the distance threshold and DT_{Min} is the minimum distance threshold.

$$DT = \max\left(\frac{S_d}{N}\sum_{t=1}^{N} DE_t, DT_{Min}\right) \tag{2.3}$$

After the best-fit plane is estimated with RANSAC, the "inliers"points are obtained. N_P is defined as the minimum number of points to fit the plane. The plane where the circular arc is located is successfully estimated if both of the following conditions are met:

- The number of the "inliers"points is larger than the threshold N_P.
- For each "inliers"point, its point-plane distance error DE_t is smaller than the distance threshold DT.

Otherwise the data to estimate the plane is considered invalid and needs to be measured again.

As shown in Figure 2.1, after estimating the normal vector $\boldsymbol{n}$, the camera coordinate frame XYZ is rotated into a new coordinate frame such that its $Y'Z'$ plane is parallel to the estimated plane and $\boldsymbol{n}$ becomes the X' axis. For the remaining "inliers"point $\boldsymbol{P}_t$, its position in the new coordinate frame is denoted as $\boldsymbol{P}'_t(\boldsymbol{P}'_t = [P'_{xt} \quad P'_{yt} \quad P'_{zt}]^T)$. The traditional and proposed methods of fitting the circular arc in 2D space are described in detail.

2.1.2 Traditional method of fitting a circular arc in 2D space

For the circular arc fitted with the remaining points after the best-fit plane estimation, in the new coordinate frame its center is set to be $\boldsymbol{C}'(\boldsymbol{C}' = [C'_x \quad C'_y \quad C'_z]^T)$.

In the 2D space, the equation of the circle is defined as (2.4) and the Euclidean (geometric) distance g_t from the data point to the fitting circular arc is defined as (2.5) [30]. The optimization problem is defined with (2.6), where M is the number of remaining points to fit the circular arc. The minimum and maximum values of the variable $C'_v(v = y, z)$ are defined to be C'^L_v and C'^H_v respectively. Similarly, the minimum and maximum values of the variable r are defined to be r^L and r^H respectively. Traditionally, The center C'_y, C'_z and radius r of the circular arc can be calculated by solving the optimization problem defined with (2.6) using Levenberg-Marquardt Algorithm (LMA) [26–28]. C'_x is solved with (2.7). After $\boldsymbol{C}'$ is obtained, $\boldsymbol{C}$

can be calculated through the rigid transform matrix between the camera coordinate frame and new coordinate frame.

$$(P'_{yt} - C'_y)^2 + (P'_{zt} - C'_z)^2 = r^2 \tag{2.4}$$

$$g_t = |\sqrt{(P'_{yt} - C'_y)^2 + (P'_{zt} - C'_z)^2} - r| \tag{2.5}$$

$$\min \sum_{t=1}^{M} {g_t}^2 \tag{2.6}$$

s.t.

$$\begin{aligned} C'^L_y \leq C'_y \leq C'^H_y \\ C'^L_z \leq C'_z \leq C'^H_z \\ r^L \leq r \leq r^H \end{aligned}$$

$$C'_x = \frac{1}{M} \sum_{t=1}^{M} P'_{xt} \tag{2.7}$$

However, when the rotation range of the manipulator's link is small, in the camera coordinate frame the trajectory of the reference point is a short arc. The traditional method does not perform well when fitting a short arc.

2.1.3 Proposed method: Fitting a circular arc in 2D space with joint angle constraint

In order to accurately fit a short arc, in this Chapter the joint angle measured by high-precision sensor (e.g. encoder) is added as a constraint to fit the circular arc in the 2D space. $\tilde{\theta}_u(t)$ denotes the joint angle measured by sensor, where u is the number of the joint which is rotated in the single-joint rotation and t is the number of the reference point position in the single-joint rotation.

Point $\boldsymbol{P}'_t$ meets the equation of the circle (2.4). Besides, as shown in Figure 2.2, for points $\boldsymbol{P}'_{t-1}$ and $\boldsymbol{P}'_t$, they nominally meet the joint angle

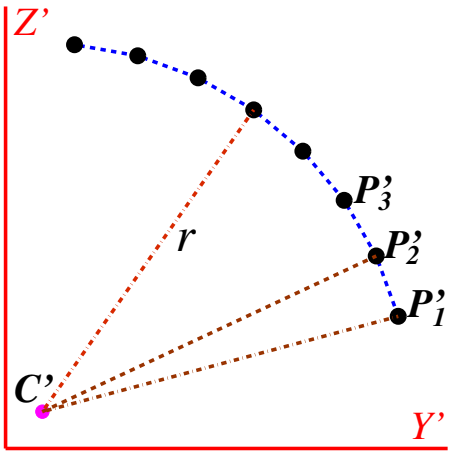


Figure 2.2: Fitting the center and radius of the circular arc in 2D space. In the new coordinate frame, for the circular arc, $\boldsymbol{C}'$ is its estimated center and r is its estimated radius.

constraint (2.8).

$$\angle \boldsymbol{P}'_{t-1}\boldsymbol{C}'\boldsymbol{P}'_t = \Delta\tilde{\theta}_u(t) = \tilde{\theta}_u(t) - \tilde{\theta}_u(t-1) \tag{2.8}$$

φ_0 is defined as the artificial initial angle of the circular arc. Since point $\boldsymbol{P}'_t$ satisfies the equation of the circle (2.4) and the joint angle constraint (2.8), its position can be defined as (2.9). The Euclidean (geometric) distances from the data point to the fitting circular arc, ${}^{y}h_t$ and ${}^{z}h_t$, are defined with (2.10).

$$\begin{cases} P'_{yt} = C'_y + r\cos(\tilde{\theta}_u(t) + \varphi_0) \\ \\ P'_{zt} = C'_z + r\sin(\tilde{\theta}_u(t) + \varphi_0) \end{cases} \tag{2.9}$$

$$\begin{cases} {}^{y}h_t = |P'_{yt} - C'_y - r\cos(\tilde{\theta}_u(t) + \varphi_0)| \\ \\ {}^{z}h_t = |P'_{zt} - C'_z - r\sin(\tilde{\theta}_u(t) + \varphi_0)| \end{cases} \tag{2.10}$$

For the circular arc, its artificial initial angle φ_0, its radius r and its center

C'_y, C'_z can be calculated through solving the optimization problem defined with (2.11).

$$\min \sum_{t=1}^{M} ({}^{y}h_t{}^2 + {}^{z}h_t{}^2) \tag{2.11}$$

s.t.

$$\begin{aligned} C'^L_y \leq C'_y \leq C'^H_y \\ C'^L_z \leq C'_z \leq C'^H_z \\ r^L \leq r \leq r^H \\ -\pi \leq \varphi_0 \leq \pi \end{aligned}$$

For point $\boldsymbol{P}'_t$, its radius error RE_t is defined with (2.12). The function $atan2(y, x)$ is used to calculate the arc tangent of y/x using the signs of arguments to determine the correct quadrant. For point $\boldsymbol{P}'_t$, its angle error AE_t is defined with (2.13).

$$RE_t = |\sqrt{(P'_{yt} - C'_y)^2 + (P'_{zt} - C'_z)^2} - r| \tag{2.12}$$

$$AE_t = |atan2((P'_{zt} - C'_z), (P'_{yt} - C'_y)) - \tilde{\theta}_u(t) - \varphi_0| \tag{2.13}$$

With RANSAC, the circular arc is estimated and its artificial initial angle φ_0, its radius r and its center C'_y, C'_z are solved. When fitting the circular arc with RANSAC, for estimating the parameters, the optimization problem defined with (2.11) is solved with LMA and the parameters φ_0, C'_y, C'_z, r are obtained. The radius error RE_t and the angle error AE_t are used to estimate the error.

The radius error threshold RT is defined with (2.14), where S_r is the scale factor of radius error threshold and RT_{Min} is the minimum radius error threshold. The angle threshold AT is defined with (2.15), where S_a is the scale factor of angle error threshold and AT_{Min} is the minimum angle error threshold.

$$RT = \max(\frac{S_r}{M} \sum_{t=1}^{M} RE_t, RT_{Min}) \tag{2.14}$$

$$AT = \max(\frac{S_a}{M}\sum_{t=1}^{M} AE_t, AT_{Min}) \tag{2.15}$$

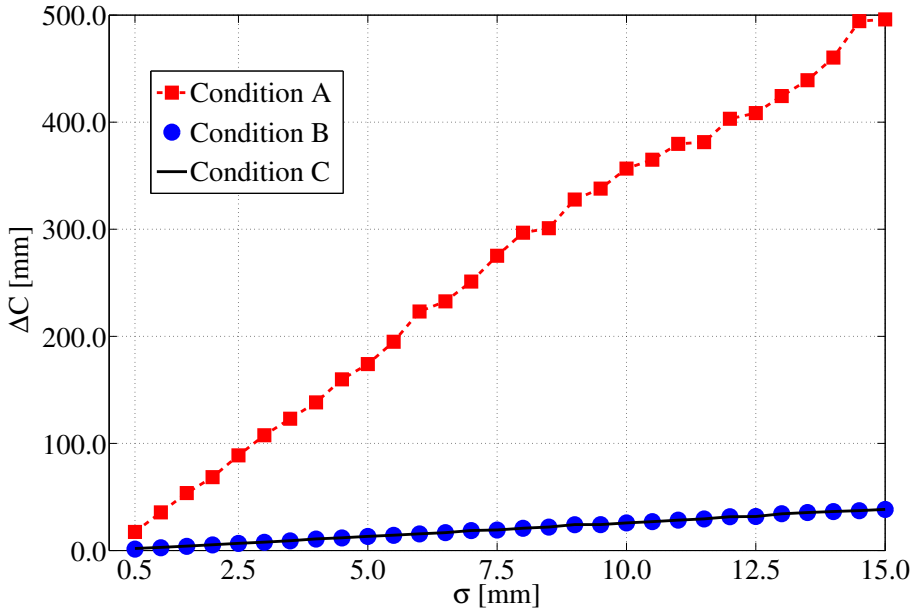
After a circular arc is fitted with RANSAC, the "inliers"points are obtained. N_C is set to be the minimum number of points to fit the circular arc. The circular arc is successfully fitted if all of the following conditions are met:

- The number of the "inliers"points is larger than the threshold N_C.
- For each "inliers"point, its radius error RE_t is smaller than the radius threshold RT.
- For each "inliers"point, its angle error AE_t is smaller than the angle threshold AT.

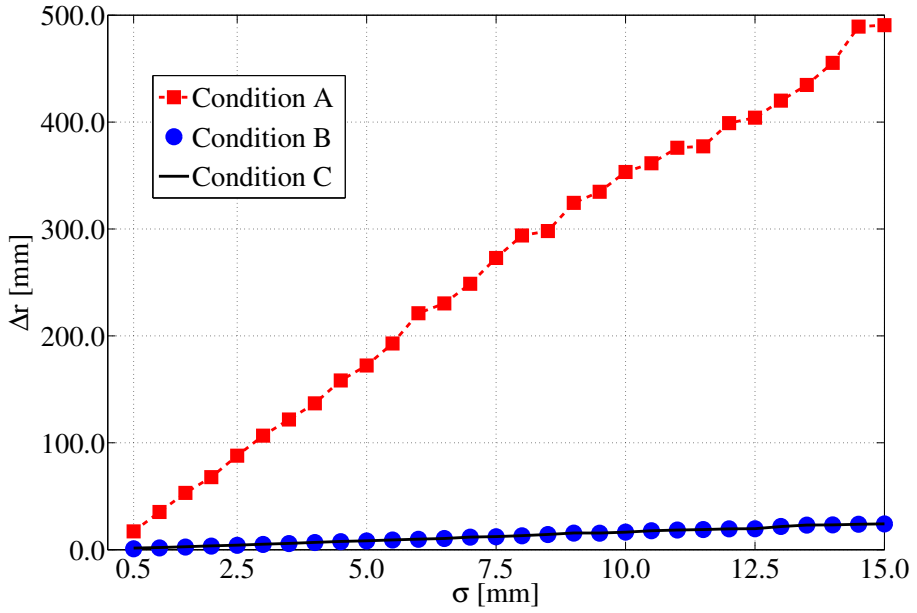
Otherwise the data to estimate the circular arc is invalid and needs to be measured again. Similarly, C'_x is solved with (2.7), and after $\boldsymbol{C}'$ is obtained, $\boldsymbol{C}$ can be calculated through the rigid transform matrix between the camera coordinate frame and new coordinate frame.

The method proposed in this Chapter was evaluated through simulation. In simulation, the center and radius of the circular arc were set to be: C'_y = 200.0 mm, C'_z = -300.0 mm, r = 2250.0 mm. It is noted that the results do not depend on the radius or the center of the circle or the location of the circular arc where the data are sampled. Tests with the circular arcs of different central angles were performed. Only the result from the circular arcs with the central angle of 20.000 deg is reported. One thousand circular arcs with the central angle of 20.000 deg were simulated. Each circular arc has 21 data points with the step angle of 1.000 deg and Gaussian noises with zero mean and different standard deviation σ were added to the positions of the data points. ΔC denotes the absolute Euclidean distance of the centers between the ground truth and the fitting result. Δr denotes the absolute radius difference between the ground truth and the fitting result. The circular arc was fitted in the following three conditions.

- *Condition A:* The circular arc was fitted with the traditional method (without joint angle constraint).



(a) The absolute Euclidean distance of the centers between the ground truth and the fitting result.



(b) The absolute radius difference between the ground truth and the fitting result.

Figure 2.3: Fitting circular arc in 2D space. One thousand circular arcs with the central angle of 20.000 deg and the step angle of 1.000 deg were simulated. Gaussian noises with zero mean and different standard deviation σ were added to the positions of the data points. The circular arc was fitted in three conditions and the averaged results from the 1000 data sets are presented.

- *Condition B:* There was no joint angle error. The circular arc was fitted with the proposed method (with joint constraint).
- *Condition C:* Gaussian noises with the zero mean and standard deviation of 0.020 deg were added to the joint angle. The circular arc was fitted with the proposed method (with joint constraint).

Figure 2.3 shows the averaged results from these 1000 data sets in three different conditions and illustrates that with the joint angle constraint the radius and center of the circular arc can be estimated more accurately. Even when Gaussian noises with zero mean and standard deviation of 0.020 deg were added to the joint angle, the circular arc can still be estimated more accurately than without joint angle constraint.

2.2 Kinematic calibration and hand-eye calibration for the humanitarian demining robot through pure rotation method with joint angle constraint

2.2.1 Description of the system

The manipulator model of the Humanitarian Demining Robot is shown in Figure 2.4. The main parameters used to model the system are summarized in Table 2.1 and listed below.

θ_{Cx}, θ_{Cy}, θ_{Cz} are the stereo vision camera fixing orientation on the first link of manipulator, $\boldsymbol{\theta}_C = [\theta_{Cx}, \theta_{Cy}, \theta_{Cz}]$.

P_{Cx}, P_{Cy}, P_{Cz} are the stereo vision camera fixing position on the first link of manipulator, $\boldsymbol{P}_C = [P_{Cx}, P_{Cy}, P_{Cz}]$.

θ_{Tx}, θ_{Ty}, θ_{Tz} are the orientation of the rigid transfer matrix from external sensor coordinate frame to robot coordinate frame.

P_{Tx}, P_{Ty}, P_{Tz} are the position of the rigid transfer matrix from external sensor coordinate frame to robot coordinate frame.

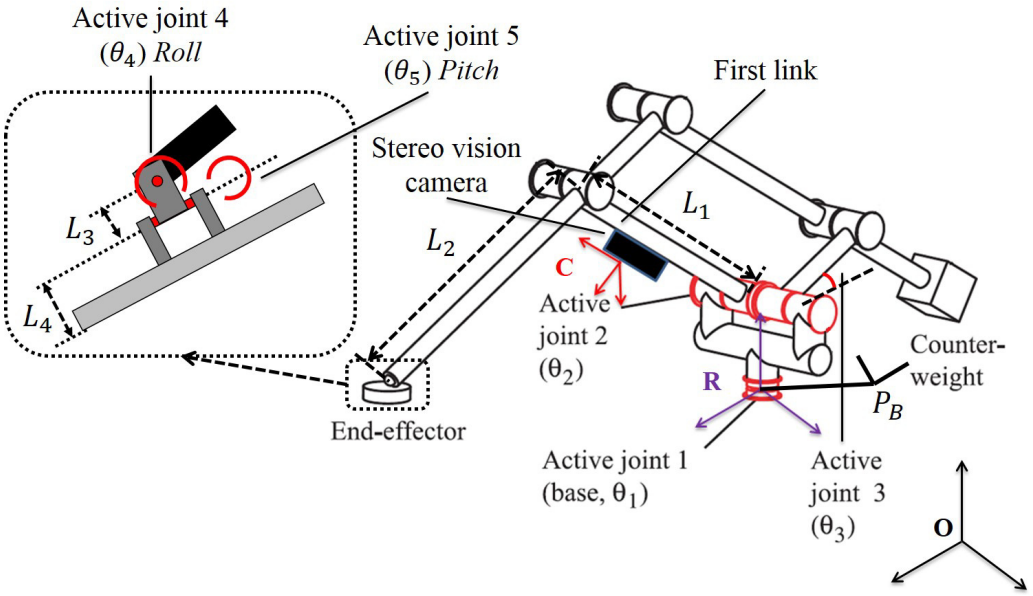


Figure 2.4: The manipulator model of the Humanitarian Demining Robot.

${}^{A}\boldsymbol{P}_{t}^{B}$ denotes that in the coordinate frame $A(A \in \{R, C, O\})$ as shown in Figure 2.4, the position of the reference point with the sequence number t, and the position is measured from $B(B \in \{SE, SC, SO\})$.

${}^{A}\Delta\boldsymbol{P}_{t}^{B,D}$ denotes that in the coordinate frame $A(A \in \{R, C, O\})$, for the reference point with the sequence number t, its position difference between its positions measured from B and $D(B, D \in \{SE, SC, SO\})$.

${}^{A}\boldsymbol{C}_{m}^{B}$ denotes that in the coordinate frame $A(A \in \{R, C, O\})$, the estimated center of the circular arc in the motion sequence $m(m = 1, 2, 3)$ and its reference point positions are measured from $B(B \in \{SE, SC, SO\})$.

$\boldsymbol{C}_{m}^{T}$ is the theoretical center of the motion sequence $m(m = 1, 2, 3)$.

${}^{A}\boldsymbol{n}_{m}^{B}$ denotes that in the coordinate frame $A(A \in \{R, C, O\})$, the estimated norm vector of the circular arc in the motion sequence $m(m = 1, 2, 3)$ and its reference point positions are measured from $B(B \in \{SE, SC, SO\})$.

$\boldsymbol{n}_{m}^{T}$ is the theoretical norm vector of the motion sequence $m(m = 1, 2, 3)$.

${}^{R}\boldsymbol{P}_{t}^{SE}$ denotes the reference point position in the robot coordinate frame and is calculated through the kinematic model with joint the angle values measured by sensors.

Table 2.1: Main parameters

Symbol	Description
$X_R Y_R Z_R(R)$	Robot coordinate frame
$X_C Y_C Z_C(C)$	Camera coordinate frame
$X_O Y_O Z_O(O)$	External sensor coordinate frame
SE	Reference point position is calculated from kinematic model with joint angles measured by sensors
SC	Reference point position is detected by stereo vision camera
SO	Reference point position is obtained by external sensor
$\tilde{\theta}_u$	Measured joint angle (with sensor) of joint $u(u = 1, 2, \ldots, 5)$
$\hat{\theta}_u$	Initial joint angle of joint $u(u = 1, 2, \ldots, 5)$
θ_u	Actual joint angle of joint $u(u = 1, 2, \ldots, 5)$, $\theta_u = \tilde{\theta}_u + \hat{\theta}_u$, $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]$
N_m	Number of the measured reference points in the motion sequence $m(m = 1, 2, 3)$

${}^{C}\boldsymbol{P}_t^{SC}$ denotes the reference point position in the camera coordinate frame and is measured by stereo vision camera.

${}^{O}\boldsymbol{P}_t^{SO}$ denotes the reference point position in the external sensor coordinate frame and is detected by external sensor.

The link lengths L_1, L_2, L_3 and L_4 are 1300.0 mm, 2105.0 mm, 25.0 mm and 76.0 mm respectively. The manufacturing error and assembly error of the manipulator are smaller than 1.0 millimeter. A stereo vision camera is attached to the first link of the manipulator and its position and orientation are defined as $(\boldsymbol{P}_C, \boldsymbol{\theta}_C) \in \boldsymbol{R}^6$. $\boldsymbol{P}_B \in \boldsymbol{R}^3$ represents the counterweight initial fixing position and defines the initial angles of joint 1, 2 and 3.

Furthermore, the manipulator has a total of 5 degrees of freedom (DOF), represented by the joint angles $\boldsymbol{\theta} \in \boldsymbol{R}^5$. It is composed of a 3-DOF pantographic manipulator with a counterweight (battery inside) that balances the total structure regardless of change in joint angles, thus reducing power consumption and ensuring the ATV's stability [24]. The joint angles of active joints 1, 2 and 3 are measured by resolvers with the resolution of 7.85×10^{-5} deg. A MMD is mounted on a 2-DOF wrist at the end of the manipulator.

2.2.2 Formulation of the calibration problem

In the robot coordinate frame, the reference point position ${}^{R}\boldsymbol{P}_t^{SC}$ is calculated with (2.16) using its position ${}^{C}\boldsymbol{P}_t^{SC}$ detected by stereo vision camera. For the reference point on the manipulator's tip, the value E_{CE} defined with (2.17) is in terms of Root Mean Square (RMS) difference between ${}^{R}\boldsymbol{P}_t^{SC}$ and ${}^{R}\boldsymbol{P}_t^{SE}$ of the entire set of N data points.

$$\begin{aligned} {}^{R}\boldsymbol{P}_t^{SC} \;=\; & R(z,\theta_1)R(x,\theta_2)(R(x,(\theta_{Cx}+\pi/2))R(y,\theta_{Cy})R(z,(\theta_{Cz}+\pi)){}^{C}\boldsymbol{P}_t^{SC} \\ & +T(P_{Cx},P_{Cy},P_{Cz})) \end{aligned} \tag{2.16}$$

$$E_{CE}=\sqrt{\frac{\sum_{t=1}^{N}\left|{}^{R}\boldsymbol{P}_t^{SC}-{}^{R}\boldsymbol{P}_t^{SE}\right|^2}{N}} \tag{2.17}$$

Table 2.2: The sensitivity of E_{CE} to parameters $\hat{\theta}_u(u=1,2,\ldots,5)$ and $(\boldsymbol{P}_C,\boldsymbol{\theta}_C)$

u	$\lvert\partial E_{CE}/\partial\hat{\theta}_u\rvert$ [mm/mrad]	v	$\lvert\partial E_{CE}/\partial\theta_{Cv}\rvert$ [mm/mrad]	$\lvert\partial E_{CE}/\partial P_{Cv}\rvert$ [mm/mm]
1	0.00	x	2.20	1.00
2	2.20	y	1.64	1.00
3	2.20	z	1.49	1.00
4	0.10			
5	0.08			

Through simulated measurements, the sensitivity of E_{CE} to the parameters $\hat{\theta}_u(u=1,2,\ldots,5)$ and $(\boldsymbol{P}_C,\boldsymbol{\theta}_C)$ is shown in Table 2.2. It is noticed that parameters $\boldsymbol{P}_C$ have nearly unity sensitivity with respect to RMS error. Most of the angular errors have balanced sensitivities with the exception of $\hat{\theta}_1$, $\hat{\theta}_4$, $\hat{\theta}_5$ and the nominal values are chosen for these parameters. Because the stereo vision camera is fixed on the first link of the manipulator, E_{CE} is not sensitive to $\hat{\theta}_1$. As L_3 and L_4 are shorter than L_1 and L_2 and they are not on the same order, $\hat{\theta}_4$ and $\hat{\theta}_5$ have less influence on E_{CE}. In summary,

the parameters to be identified in the kinematic calibration and hand-eye calibration are chosen as $\boldsymbol{p} = \{\hat{\theta}_2, \hat{\theta}_3, \theta_{Cx}, \theta_{Cy}, \theta_{Cz}, P_{Cx}, P_{Cy}, P_{Cz}\}$.

2.2.3 Calibration using pure rotation method with joint angle constraint method

In the camera coordinate frame, when the stereo vision camera views a reference point on the ground it is a 2-DOF system and when it views at a reference point on manipulator's tip with active joint 4 and active joint 5 fixed and it is only 1-DOF system. In order to obtain all the parameters $\boldsymbol{p}$, two additional reference points are needed: a fixed internal reference point on manipulator's tip, and another external arbitrary reference point fixed in the robot coordinate frame. With stereo vision camera, the position of the reference point can be detected in the camera coordinate. Hence, the proposed calibration method, Pure Rotation with Joint Angle Constraint Method, is applied to the Humanitarian Demining Robot with the following steps.

- *Step 1.* Data acquisition.

  For each motion sequence m, the reference point position ${}^{C}\boldsymbol{P}_{m,t}^{SC}$ is measured by stereo vision camera and the joint angles $\tilde{\theta}_u(m,t)$ are measured with the high-precision resolvers of the Humanitarian Demining Robot system.

  - *Step 1.1.* Trajectory of internal reference point.

    Rotate the active joint 3 with $N_1 (N_1 \geq 3)$ steps, while keeping all other active joints unchanged. The stereo vision camera views the fixed internal reference point on manipulator's tip. The trajectory of the reference point measured by stereo vision camera is defined as motion sequence 1 and described with points ${}^{C}\boldsymbol{P}_{1,t}^{SC}, t = 1, 2, \ldots, N_1$.
  - *Step 1.2.* Trajectories of external reference point.

    Rotate the active joint 2 and active joint 1 one at a time with $N_m (N_m \geq 3, m = 2, 3)$ steps, while freezing all other active joints.

The camera views an external arbitrary reference point fixed in the robot coordinate frame. The trajectory of the reference point measured by stereo vision camera is called motion sequence m and described with points ${}^{C}\boldsymbol{P}_{m,t}^{SC}, t = 1, 2, \ldots, N_m$.

- *Step 2.* Parameters calculation.

  In the camera coordinate frame, for each motion sequence $m(m = 1, 2, 3)$, using its data set of the reference point positions ${}^{C}\boldsymbol{P}_{m,t}^{SC}$ and joint angles $\tilde{\theta}_u(m, t)$, with the proposed Pure Rotation with Joint Angle Constraint Method, the normal vector ${}^{C}\boldsymbol{n}_m^{SC}$ and center ${}^{C}\boldsymbol{C}_m^{SC}$ of the circular arc are solved. Since the joint axis of the moved joint must pass through the center of this circular arc, from the mechanical structure, the theoretical rotation axis $\boldsymbol{n}_m^T$ and theoretical rotation center $\boldsymbol{C}_m^T$ of the motion sequence m are known. For the motion sequence m, it meets (2.18) and (2.19). $G_m(\boldsymbol{p}, {}^{C}\boldsymbol{n}_m^{SC})$ represents the relationship between $\boldsymbol{n}_m^T$ and ${}^{C}\boldsymbol{n}_m^{SC}$. $F_m(\boldsymbol{p}, {}^{C}\boldsymbol{C}_m^{SC})$ represents the relationship between $\boldsymbol{C}_m^T$ and ${}^{C}\boldsymbol{C}_m^{SC}$. $G_m(\boldsymbol{p}, {}^{C}\boldsymbol{n}_m^{SC})$ and $F_m(\boldsymbol{p}, {}^{C}\boldsymbol{C}_m^{SC})$ can be obtained through the geometry principles of robot configuration. After the normal vectors and centers of the circular arcs in the motion sequence 1, 2 and 3 are estimated, the calibration parameters $\boldsymbol{p} = \{\hat{\theta}_2, \hat{\theta}_3, \theta_{Cx}, \theta_{Cy}, \theta_{Cz}, P_{Cx}, P_{Cy}, P_{Cz}\}$ can be solved. In Appendix B, this process is described in detail.

$$\boldsymbol{n}_m^T = G_m(\boldsymbol{p}, {}^{C}\boldsymbol{n}_m^{SC}) \tag{2.18}$$

$$\boldsymbol{C}_m^T = F_m(\boldsymbol{p}, {}^{C}\boldsymbol{C}_m^{SC}) \tag{2.19}$$

2.3 Experimental results

In this section, the proposed calibration method, Pure Rotation with Joint Angle Constraint Method, was applied to the Humanitarian Demining

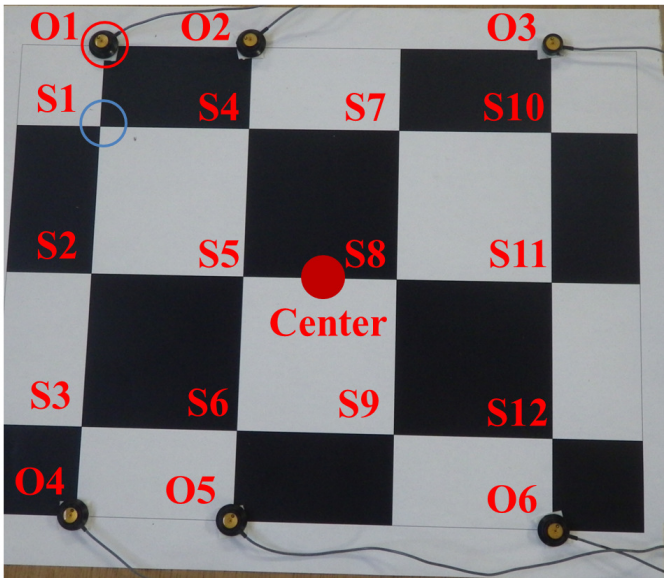


Figure 2.5: Chessboard with the squares of the same size was used in this Chapter. In the camera coordinate frame, the positions of the corners $S1$ to $S12$ were measured with stereo vision camera, and the center ${}^{C}\boldsymbol{P}^{SC}$ was calculated with the positions of $S1$ to $S12$. The markers of the Optotrak 3020 were attached to the corners of chessboard. In the Optotrak coordinate frame, the positions of the markers $O1$ to $O6$ were detected with Optotrak 3020, and the center ${}^{O}\boldsymbol{P}^{SO}$ of the chessboard was calculated with the positions of $O1$ to $O6$.

Robot. It was verified and compared with other commonly used calibration methods.

2.3.1 Calibration using stereo vision camera

The Point Grey Research's *Bumblebee*^{TM} stereo vision camera, which has a rigid aluminum body to resist shocks and vibrations, was used in this Chapter. The focal length of its lenses is 6.0 mm and its Horizontal Field of View (HFOV) is 50.0 deg. It is pre-calibrated for lens distortions and camera misalignments. Besides, it doesn't require in-field calibration and is guaranteed to stay calibrated. The chessboard shown in Figure 2.5 was used in the experiment and its center was considered as the reference point. In the camera coordinate frame, the positions of its corners $S1$ to $S12$ can be

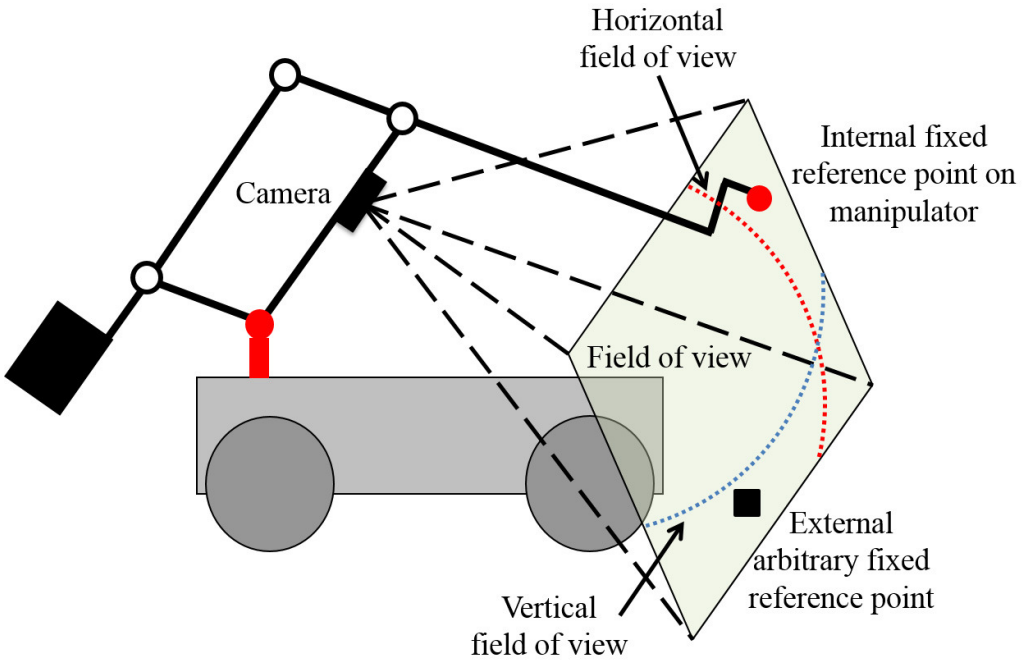


Figure 2.6: The view range of stereo vision camera.

measured with the $Bumblebee^{TM}$ stereo vision camera and its center ${}^{C}\boldsymbol{P}^{SC}$ is calculated with the positions of $S1$ to $S12$. The chessboard was attached to the manipulator's tip and the internal fixed reference point was obtained. It was put to an arbitrary position fixed in the robot coordinate frame and the external arbitrary fixed reference point was obtained. With the data acquirement method described in Section 2.2.3, the $Bumblebee^{TM}$ stereo vision camera was used to acquire the positions of the internal and external reference points. As shown in Figure 2.6, due to the limitation of the camera view range, to ensure the chessboard is viewed by stereo vision camera in each motion sequence, the motion range of the active joint that rotates in the single-joint rotation is quite limited. The joint motion range and the number of measured points in each motion sequence are shown in Table 2.3. The calibration parameters $\boldsymbol{p}$ were solved with the following methods.

- *Method A*. Calibration with the proposed method (Pure Rotation with Joint Angle Constraint Method), which is described in Section 2.1 and

Section 2.2.

- *Method B.* Calibration with pure rotation using traditional method. As described in Section 2.1 the joint angle constraint was not used to fit the circular arc and the calibration parameters were solved with the parameters calculation method described in Section 2.2.
- *Method C.* Calibration with LMA [7], which is a commonly used calibration method. For the internal reference point in motion sequence 1, its position differences are defined with (2.20). In the robot coordinate frame, since the external reference point is fixed, with its positions measured in motion sequence 2 and 3, its averaged position $\overline{{}^{R}\boldsymbol{P}_{ext}^{SC}}$ can be estimated with (2.21). For the external reference point in motion sequence 2 and 3, its position differences are defined with (2.22). The minimum and maximum values of the variable $\hat{\theta}_u (u = 2, 3)$ are defined to be θ_u^L and θ_u^H respectively. Similarly, the minimum and maximum values of the variable $\theta_{Cv} (v = x, y, z)$ are defined to be θ_{Cv}^L and θ_{Cv}^H respectively. The minimum and maximum values of the variable $P_{Cv} (v = x, y, z)$ are defined to be P_{Cv}^L and P_{Cv}^H respectively. The optimization problem defined with (2.23) was solved with LMA and the calibration parameters $\boldsymbol{p}$ were calculated.

$$E_{int} = \frac{1}{N_1} \sum_{t=1}^{N_1} \left| {}^{R}\boldsymbol{P}_{1,t}^{SC} - {}^{R}\boldsymbol{P}_{1,t}^{SE} \right|^2 \tag{2.20}$$

$$\overline{{}^{R}\boldsymbol{P}_{ext}^{SC}} = \frac{1}{\sum_{m=2}^{3} N_m} \sum_{m=2}^{3} \sum_{t=1}^{N_m} {}^{R}\boldsymbol{P}_{m,t}^{SC} \tag{2.21}$$

$$E_{ext} = \frac{1}{\sum_{m=2}^{3} N_m} \sum_{m=2}^{3} \sum_{t=1}^{N_m} \left| {}^{R}\boldsymbol{P}_{m,t}^{SC} - \overline{{}^{R}\boldsymbol{P}_{ext}^{SC}} \right|^2 \tag{2.22}$$

$$\min \left(E_{int} + E_{ext} \right) \tag{2.23}$$

s.t.

$$\theta_u^L \leq \hat{\theta}_u \leq \theta_u^H, \qquad u = 2, 3$$

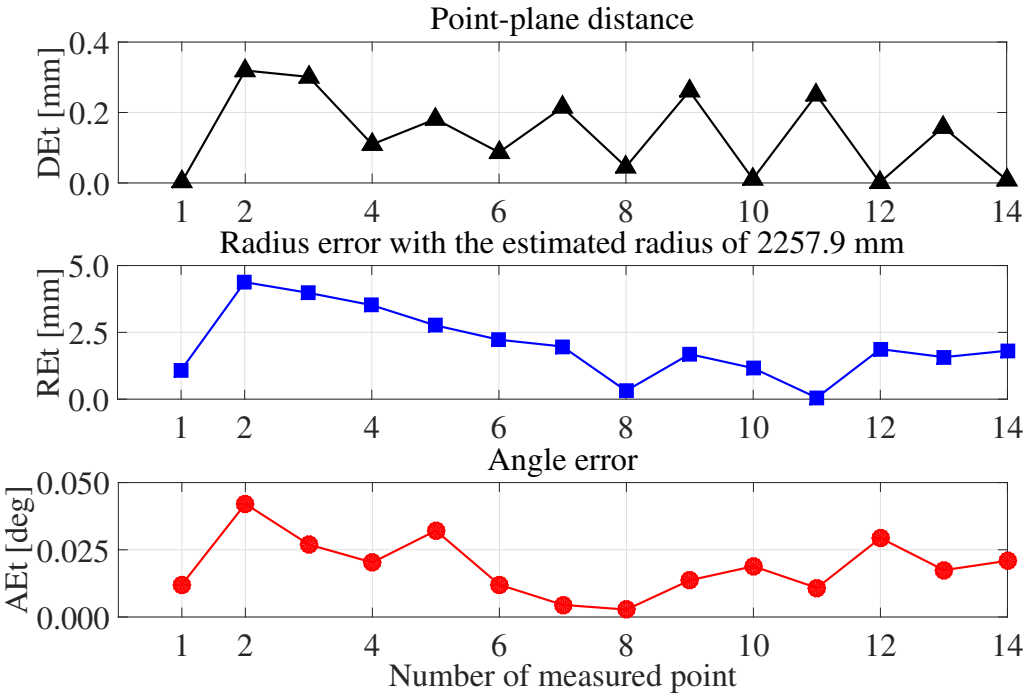


Figure 2.7: A circular arc measured in experiment was fitted with the proposed method (Pure Rotation with Joint Angle Constraint Method) and its point-plane distance error DE_t, its radius error RE_t and its angle error AE_t are presented. The parameters: $N = 15$, $S_d = 2.0$, $DT_{Min} = 0.5$ mm, $N_P = 13$, $S_r = 2.0$, $RT_{Min} = 5.0$ mm, $S_a = 2.0$, $AT_{Min} = 0.050$ deg, $N_C = 13$.

$$\theta_{Cv}^L \leq \theta_{Cv} \leq \theta_{Cv}^H, \qquad v = x, y, z$$

$$P_{Cv}^L \leq P_{Cv} \leq P_{Cv}^H, \qquad v = x, y, z$$

The performance of the proposed calibration method was evaluated with data measured in experiment. A circular arc measured in experiment was fitted with the proposed method and its point-plane distance error DE_t, its radius error RE_t and its angle error AE_t are shown in Figure 2.7.

The point-plane distance error DE_t is defined with (2.2) and the RMS of point-plane distance error D_S is defined as (2.24). The radius error RE_t is defined with (2.12) and the RMS of radius error R_S is defined as (2.25). The angle error AE_t is defined with (2.13) and the RMS of angle error A_S is

Table 2.3: The experimental settings of the calibration and the calibration results of the proposed method (Pure Rotation with Joint Angle Constraint Method)

	Sequence 1	Sequence 2	Sequence 3
Number of measured points	15	15	15
Joint motion range [deg]	18.204	24.498	34.968
Estimated radius [mm]	1989.7	2257.9	1990.8
RMS of DE_t [mm]	0.1	0.2	5.7
RMS of RE_t [mm]	1.5	2.4	4.5
RMS of AE_t [deg]	0.017	0.036	0.013

Table 2.4: With pure rotation, the theoretical centers and estimated centers of the circular arc in each motion sequence

	Theoretical value [mm]			Proposed method [mm]		
	C_x	C_y	C_z	C_x	C_y	C_z
Sequence 1	#	0.0	1300.0	#	0.0	1283.1
Sequence 2	#	0.0	0.0	#	0.0	16.9
Sequence 3	0.0	0.0	#	-0.5	12.3	#
	Traditional method [mm]					
	C_x	C_y	C_z			
Sequence 1	#	0.0	1271.1			
Sequence 2	#	0.0	28.9			
Sequence 3	-0.6	-31.4	#			

Note:
# denotes the value unknown, hereafter it refers to the same meaning.

defined as (2.26).

$$D_S = \sqrt{\frac{1}{N}\sum_{t=1}^{N} {DE_t}^2} \tag{2.24}$$

$$R_S = \sqrt{\frac{1}{N}\sum_{t=1}^{N} {RE_t}^2} \tag{2.25}$$

$$A_S = \sqrt{\frac{1}{N}\sum_{t=1}^{N} {AE_t}^2} \tag{2.26}$$

After fitting the circular arcs with the proposed calibration method, for each motion sequence, the estimated radius of the circular arc, the RMS of point-plane distance error D_S, the RMS of radius error R_S and the RMS of angle error A_S are summarized in Table 2.3. The theoretical centers and estimated centers of the circular arc in each motion sequence are summarized in Table 2.4. The estimated centers were obtained with the proposed calibration method and traditional calibration method. From this table, it can be concluded that with the joint angle constraint the center of the circular arc can be estimated more accurately even though the joint motion range is small and the radius of the circular arc is large as shown in Table 2.3.

2.3.2 Verifying and comparing the calibration results

Optotrak 3020 (Northern Digital Inc.) was used as an external hardware to verify the calibration results solved in Section 2.3.1 due to its high precision, repeatability and accuracy [23]. As reported by the manufacturer, the RMS accuracy of the system at 2.25 m distance was 0.1 mm for x, y coordinates and 0.15 mm for z coordinate.

A. Results of the kinematic calibration

In order to check the result of kinematic calibration, the marker of Optotrak 3020 was attached to manipulator' tip (Figure 2.8) and it was considered as the reference point. In the robot coordinate frame, for the marker on the manipulator's tip, its trajectories of rectangle motion and circle motion were generated through inverse kinematics with setting the joint initial angles to be the nominal value. As shown in Figure 2.9, the rectangle motion with the area of 1000.0 mm x 2000.0 mm at different heights with the increment of 100.0 mm was done and described with points ${}^{O}\boldsymbol{P}_{4,t}^{SO}, t = 1, 2, \ldots, N_4$ (N_4 is 264 in this Chapter). As presented in Figure 2.10, the circle motion with the diameter of 1000.0 mm at different heights with the increment of 100.0 mm

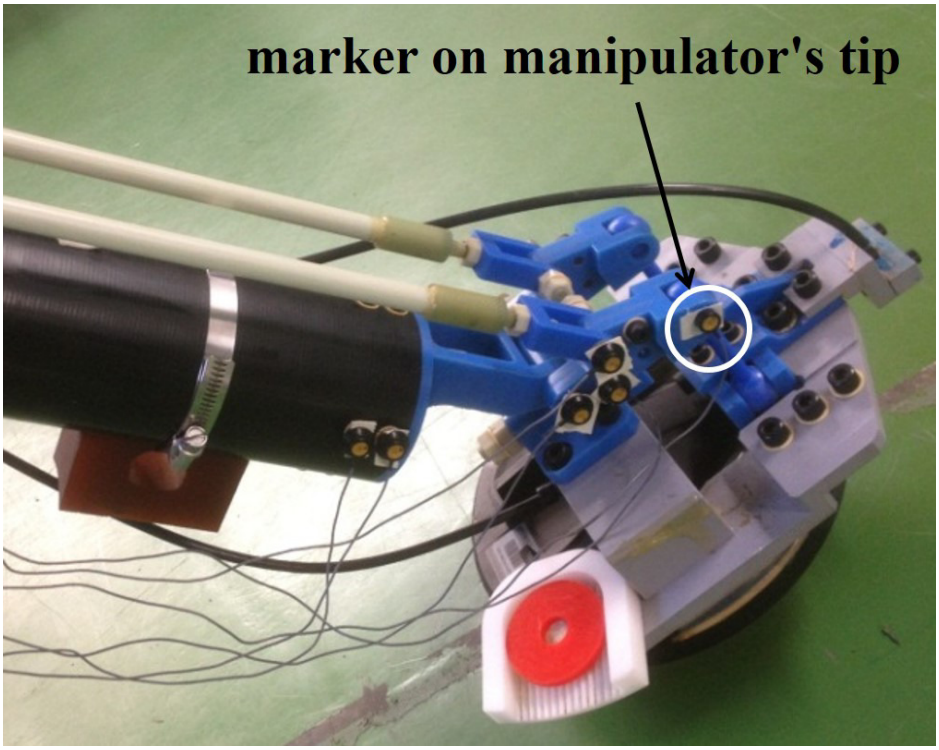


Figure 2.8: Manipulator's tip. The markers of Optotrak 3020 were attached to the manipulator.

was done and described with points ${}^{O}\boldsymbol{P}_{5,t}^{SO}, t = 1, 2, \ldots, N_5$ (N_5 is 160 in this Chapter). The joint initial angles $\hat{\theta}_2$ and $\hat{\theta}_3$ were set to the following values solved in Section 2.3.1.

- *Before calibration.* Nominal value.
- *Method A.* Calibration with the proposed method (Pure Rotation with Joint Angle Constraint Method).
- *Method B.* Calibration with pure rotation using traditional method (the joint angle constraint is not used to fit the circular arc).
- *Method C.* Calibration with LMA.

In the robot coordinate frame, the reference point position ${}^{R}\boldsymbol{P}_{t}^{SO}$ is calculated with (2.27) using its position ${}^{O}\boldsymbol{P}_{t}^{SO}$ measured by Optotrak 3020 and

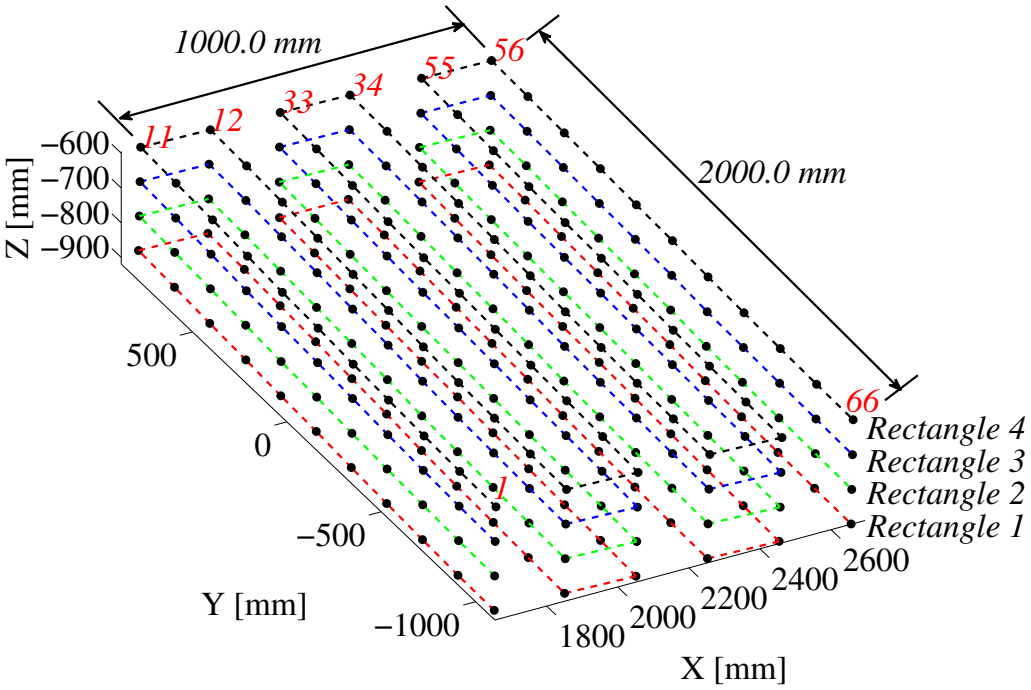


Figure 2.9: Rectangle motion. Four rectangles with the area of 1000.0 mm x 2000.0 mm were at different heights with the increment of 100.0 mm. For each rectangle, the grid point number was defined same as the Rectangle 4.

the reference point position error $|{}^{R}\Delta \boldsymbol{P}_t^{SE,SO}|$ is defined with (2.28).

$$ {}^{R}\boldsymbol{P}_t^{SO} = R(x,\theta_{Tx})R(y,\theta_{Ty})R(z,\theta_{Tz}){}^{O}\boldsymbol{P}_t^{SO} + T(P_{Tx},P_{Ty},P_{Tz}) \tag{2.27} $$

$$ |{}^{R}\Delta \boldsymbol{P}_t^{SE,SO}| = |{}^{R}\boldsymbol{P}_t^{SE} - {}^{R}\boldsymbol{P}_t^{SO}| \tag{2.28} $$

For the circle motion and rectangle motion, the reference point position errors $|{}^{R}\Delta \boldsymbol{P}_t^{SE,SO}|$ are shown in Figure 2.11. From the experimental results, it is noticed *Method A* performs better than *Method B* and *Method C*. Using the proposed method, the joint initial angles $\hat{\theta}_2$ and $\hat{\theta}_3$ were well calibrated with the reference point position errors at the manipulator's tip smaller than 11.0 mm which meet the error requirement of the Humanitarian Demining Robot system.

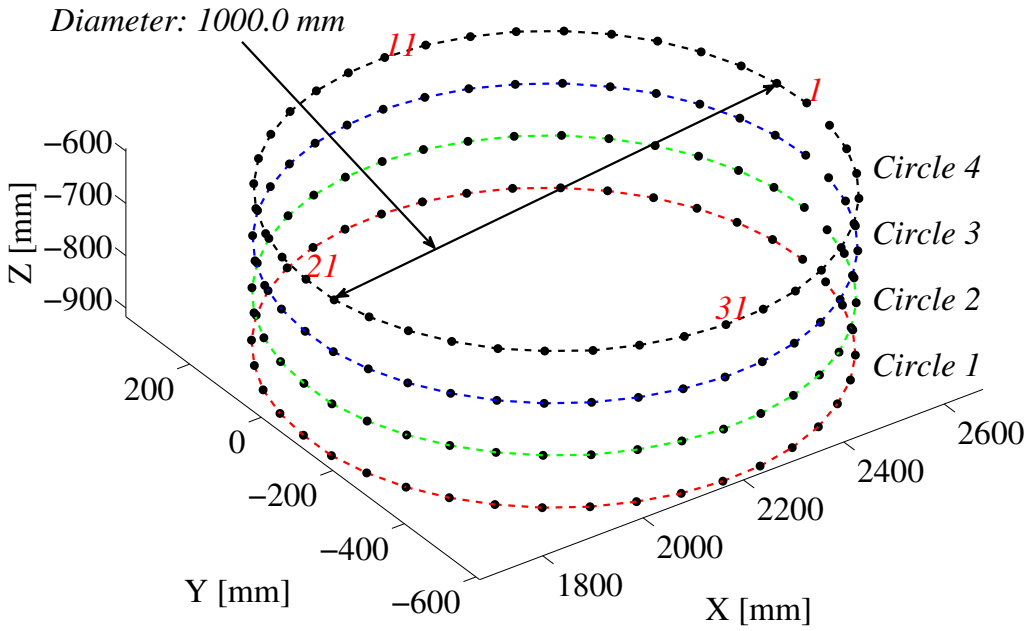
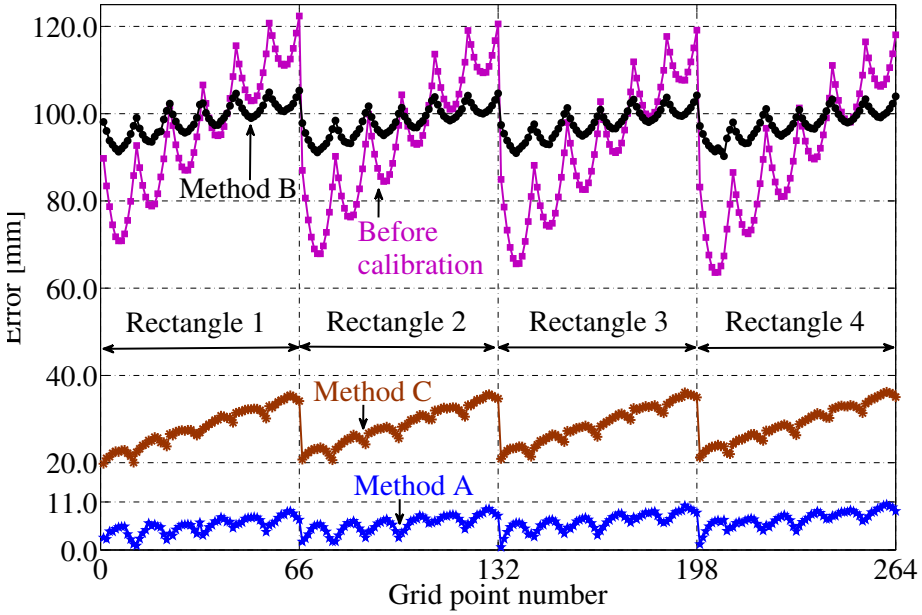


Figure 2.10: Circle motion. Four circles with the diameter of 1000.0 mm were at different heights with the increment of 100.0 mm. For each circle, the grid point number was defined same as the Circle 4.

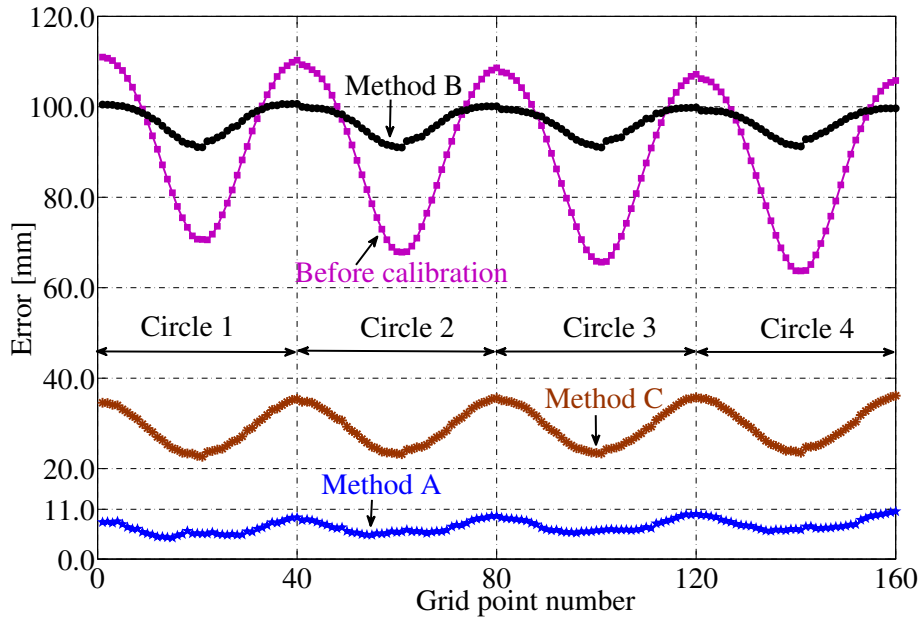
B. Comparing with the calibration result of an external high-precision calibration tool

To further verify the calibration method described in this Chapter, in laboratory, the Optotrak 3020 was used as an external high-precision calibration tool to calibrate the joint initial angles $\hat{\theta}_2$ and $\hat{\theta}_3$. The joint initial angles $\hat{\theta}_2$ and $\hat{\theta}_3$ were set to the following values.

- *Method A.* Using stereo vision camera, calibration with the proposed method (Pure Rotation with Joint Angle Constraint Method).
- *Method D.* Using Optotrak 3020, calibration with LMA. Using the data acquired in Section 2.3.2 (the rectangle motion shown in Figure 2.9 and the circle motion shown in Figure 2.10), the joint initial angles $\hat{\theta}_2$ and $\hat{\theta}_3$ were obtained by solving the optimization problem defined with (2.29)

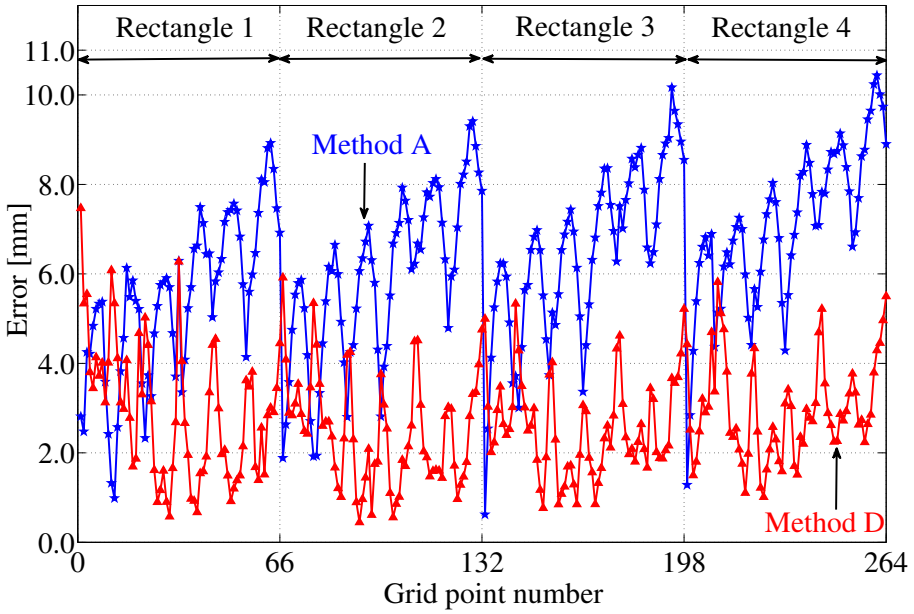


(a) The reference point position errors $|{}^{R}\Delta \boldsymbol{P}_t^{SE,SO}|$ at different positions in the rectangle motion.

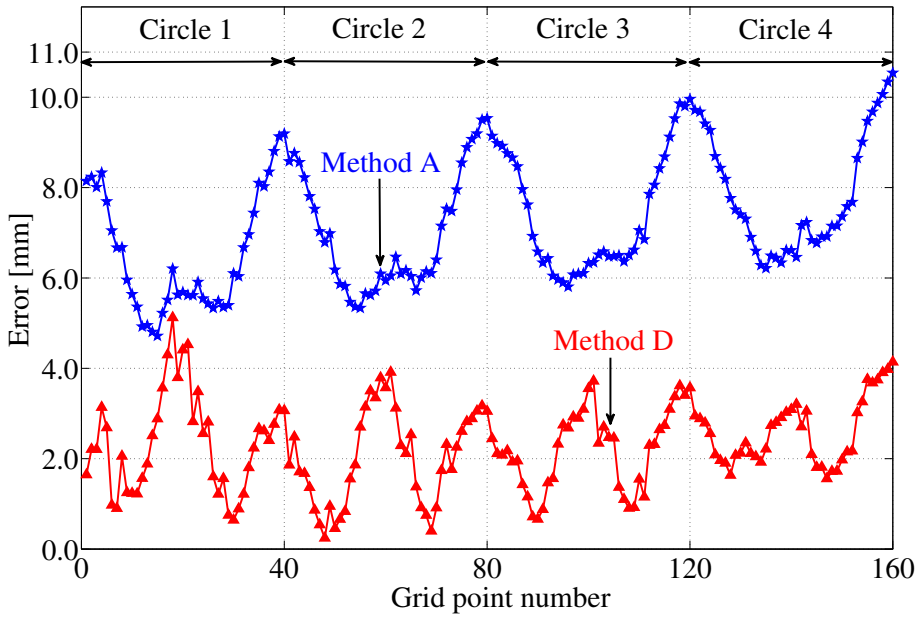


(b) The reference point position errors $|{}^{R}\Delta \boldsymbol{P}_t^{SE,SO}|$ at different positions in the circle motion.

Figure 2.11: The reference point position errors caused by the joint initial angles $\hat{\theta}_2$ and $\hat{\theta}_3$. Before calibration, nominal value. Method A, calibration with the proposed method (Pure Rotation with Joint Angle Constraint Method). Method B, calibration with pure rotation using traditional method (the joint angle constraint is not used to fit the circular arc). Method C, calibration with LMA.



(a) The reference point position errors $|{}^{R}\Delta \boldsymbol{P}_t^{SE,SO}|$ at different positions in the rectangle motion.



(b) The reference point position errors $|{}^{R}\Delta \boldsymbol{P}_t^{SE,SO}|$ at different positions in the circle motion.

Figure 2.12: The reference point position errors caused by the joint initial angles $\hat{\theta}_2$ and $\hat{\theta}_3$ when comparing with the calibration result of an external high-precision calibration tool. Method A, using stereo vision camera, calibration with the proposed method (Pure Rotation with Joint Angle Constraint Method). Method D, using Optotrak 3020, calibration with LMA.

using LMA.

$$\min\left(\sum_{t=1}^{N_4} |{}^R\boldsymbol{P}_{4,t}^{SE} - {}^R\boldsymbol{P}_{4,t}^{SO}|^2 + \sum_{t=1}^{N_5} |{}^R\boldsymbol{P}_{5,t}^{SE} - {}^R\boldsymbol{P}_{5,t}^{SO}|^2\right) \tag{2.29}$$

s.t.

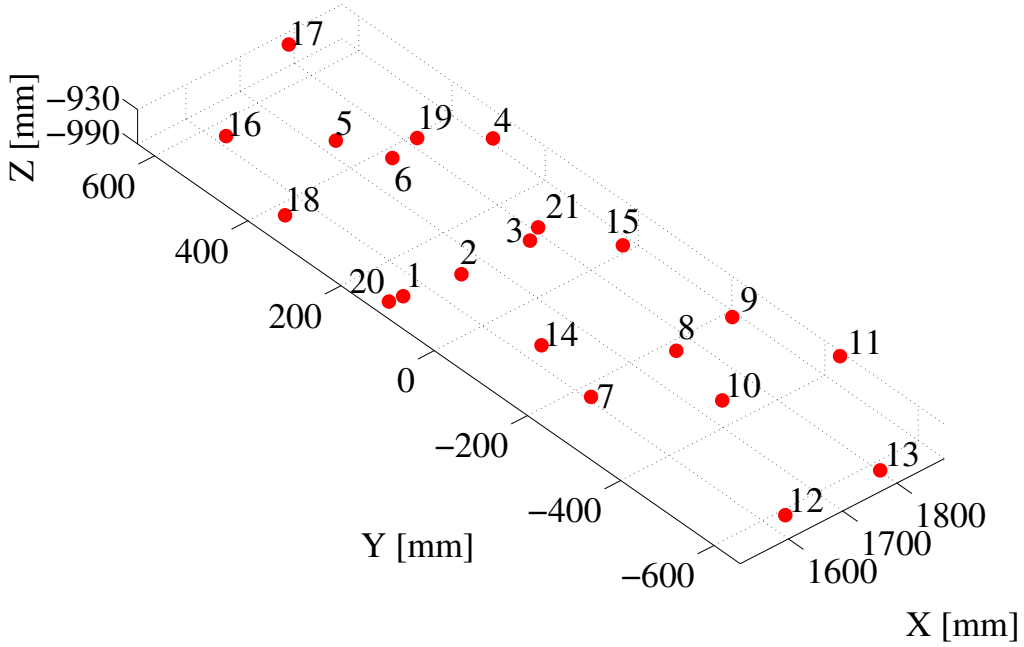
$$\theta_u^L \leq \hat{\theta}_u \leq \theta_u^H, \qquad u = 2, 3$$

With the data acquired in Section 2.3.2, for the rectangle motion shown in Figure 2.9 and circle motion shown in Figure 2.10, in the robot coordinate frame the reference point position errors $|{}^R\Delta\boldsymbol{P}_t^{SE,SO}|$ defined with (2.28) are shown in Figure 2.12. Since the external high-precision calibration tool was used in *Method D* to acquire the calibration data in the space of 1000.0 mm x 2000.0 mm x 300.0 mm, the calibration result of *Method D* is better than the result of *Method A*. However it is not practical to calibrate the Humanitarian Demining Robot system with Optotrak 3020 in field. With the proposed method the reference point position errors at the manipulator's tip were smaller than 11.0 mm and quite close to the calibration result of *Method D*.

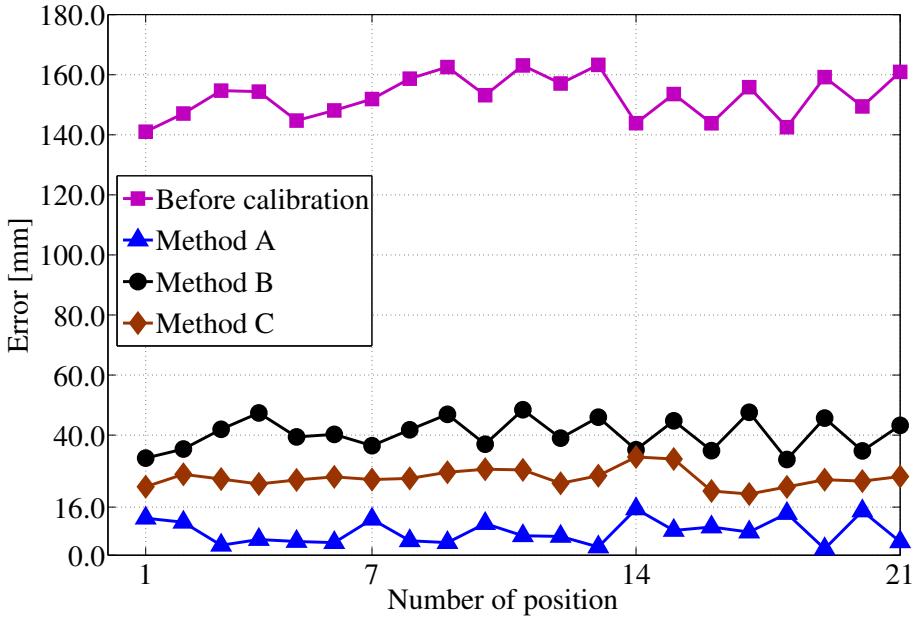
C. Results of the kinematic calibration and hand-eye calibration

In order to check calibration result of the kinematic calibration and hand-eye calibration, the center of the chessboard (Figure 2.5) is considered as the reference point. As shown in Figure 2.13(a), the chessboard was placed to 21 different random positions on ground within the area of 500.0 mm x 1200.0 mm. Because of the limited view range of the $Bumblebee^{TM}$ stereo vision camera, it was needed to rotate the manipulator to 3 different poses to ensure the stereo vision camera can view the chessboard at these 21 random positions. In the Optotrak coordinate frame, the positions of markers $O1$ to $O6$ can be accurately detected by Optotrak 3020 and the chessboard's center ${}^O\boldsymbol{P}_t^{SO}$ was calculated with the positions of $O1$ to $O6$. The calibration parameters $\boldsymbol{p}$ were set to the following values solved in Section 2.3.1.

- *Before calibration.* Nominal value.



(a) The chessboard was placed to 21 random positions on the ground.



(b) The reference point position errors $|{}^{R}\Delta \boldsymbol{P}_t^{SC,SO}|$ at different positions. Before calibration, nominal value. Method A, calibration with the proposed method (Pure Rotation with Joint Angle Constraint Method). Method B, calibration with pure rotation using traditional method (the joint angle constraint is not used to fit the circular arc). Method C, calibration with LMA.

Figure 2.13: Checking the reference point position errors caused by joint initial angles and camera fixing position and orientation.

- *Method A.* Calibration with the proposed method (Pure Rotation with Joint Angle Constraint Method).
- *Method B.* Calibration with pure rotation using traditional method (the joint angle constraint is not used to fit the circular arc).
- *Method C.* Calibration with LMA.

In the robot coordinate frame, the reference point position error $|{}^{R}\Delta \boldsymbol{P}_t^{SC,SO}|$ is defined with (2.30) and presented in Figure 2.13(b). It illustrates *Method A* performs better than *Method B* and *Method C*. After calibration with the proposed method, the stereo vision camera can map the terrain with the position errors smaller than 16.0 mm which meet the error requirement of the Humanitarian Demining Robot system.

$$|{}^{R}\Delta \boldsymbol{P}_t^{SC,SO}| = |{}^{R}\boldsymbol{P}_t^{SC} - {}^{R}\boldsymbol{P}_t^{SO}| \tag{2.30}$$

2.3.3 Discussion

Through experiments, it is verified that *Method A* proposed in this Chapter has better performance than *Method B* and *Method C* for the reason that as shown in Table 2.4 when the joint angle is added as a constraint the circular arc can be fitted more accurately. The calibration result of the proposed calibration method is quite close to the calibration result of the external high-precision calibration tool (Optotrak 3020) even though the joint motion range in each motion sequence is limited and the accuracy of the stereo vision camera used in the experiments is lower than the external high-precision calibration tool (Optotrak 3020).

2.4 Conclusion

With the proposed method (Pure Rotation with Joint Angle Constraint Method), the kinematic calibration and hand-eye calibration are performed simultaneously. The proposed calibration method was compared and verified with other calibration methods by experiments using an external high-precision hardware. Through the experiments, it is verified that the proposed

calibration method performs better than the traditional pure rotation calibration method (the joint angle constraint is not used to fit the circular arc) and the commonly used calibration method (nonlinear least squares estimation with LMA) because with the proposed calibration method a short circular arc could be fitted more accurately and the actual position of the robot's articulation axis could be accurately determined for the reason that the joint angle is added as a constraint when fitting the circular arc. With the proposed calibration method, the joint initial angles of the manipulator are calibrated with the reference point position error at the manipulator's tip smaller than 11.0 mm and the stereo vision camera can map the terrain with the position error smaller than 16.0 mm which meet the error requirement of the Humanitarian Demining Robot system. The calibration method described in this Chapter can be used to other robot platforms to do kinematic calibration and hand-eye calibration.

Chapter 3

Hand-Eye Calibration with Pure Rotation Method through Directly Fitting Circular Arc in 3D Space with Joint Angle Constraint

In this Chapter, the Pure Rotation Calibration Method described in Chapter 2 is improved and the circular arc is directly fitted in 3D space with joint angle constraint. The Chapter is organized as follows: in Section 3.1 the traditional methods of fitting the circular arc in 3D space are introduced and the proposed method of directly fitting the circular arc in 3D space is described in detail. The simulation results are shown in Section 3.2. In Section 3.3, for a Pan-Tilt-Camera system, the Hand-Eye Calibration with the proposed method is introduced and the experimental results are shown. Finally, a conclusion is given in Section 3.4.

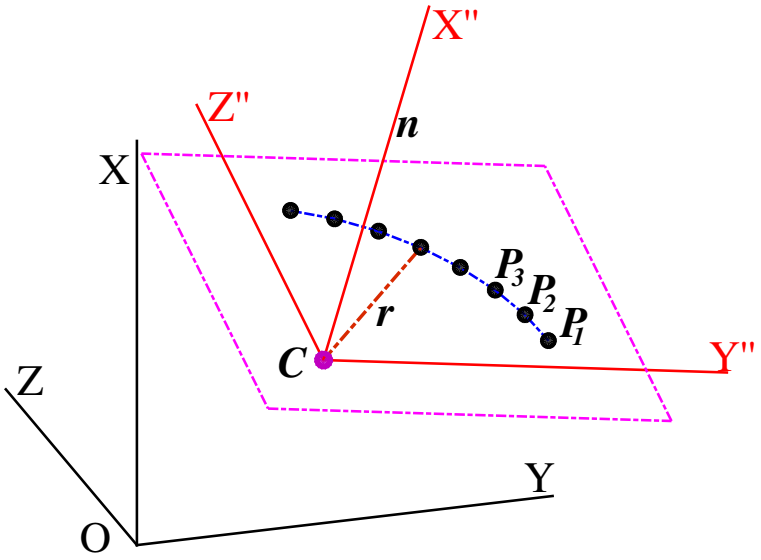


Figure 3.1: Circular arc estimation in 3D space. In the XYZ coordinate frame, for the circular arc, $\boldsymbol{n}$ is its estimated normal vector, $\boldsymbol{C}$ is its estimated center and r is its estimated radius.

3.1 Directly fitting a circular arc in 3D space

As shown in Figure 3.1, in the 3D space, given $N(N \geq 3)$ points $\boldsymbol{P}_i$, $i = 1, 2, \ldots, N$, which are nominally located in a plane that can be defined by a normal vector $\boldsymbol{n}$ and a signed distance d from the origin, a circular arc is constructed and defined by three parameters: normal vector $\boldsymbol{n}$, center $\boldsymbol{C}$ and radius r.

Note: $\boldsymbol{P}_i = [P_{xi} \quad P_{yi} \quad P_{zi}]^T$, the Euclidean norm $|\boldsymbol{P}| = \sqrt{{P_x}^2 + {P_y}^2 + {P_z}^2}$, $\boldsymbol{C} = [C_x \quad C_y \quad C_z]^T$, $\boldsymbol{n} = [n_x \quad n_y \quad n_z]^T$ with $\boldsymbol{n}^T\boldsymbol{n} = 1$.

3.1.1 Traditional method of directly fitting a circular arc in 3D space

As shown in Figure 3.1, the normal vector $\boldsymbol{n}$ can be defined with (3.1) through the variables ϕ_y and ϕ_z, where ϕ_y is the angle between $\boldsymbol{n}$ and the XY plane and ϕ_z is the angle between the projection of $\boldsymbol{n}$ onto the XY plane

and the X axis [31]. The XYZ coordinate frame is transformed into the $X''Y''Z''$ coordinate frame. In the $X''Y''Z''$ coordinate frame, the position of point $\boldsymbol{P}_i$ is defined as $\boldsymbol{P}_i''(\boldsymbol{P}_i'' = [P_{xi}'' \quad P_{yi}'' \quad P_{zi}'']^T)$ and can be calculated with (3.2).

$$\begin{cases} n_x = \cos(\phi_y)\cos(\phi_z) \\ n_y = -\cos(\phi_y)\sin(\phi_z) \\ n_z = \sin(\phi_y) \end{cases} \tag{3.1}$$

$$\boldsymbol{P}_i'' = R(y, \phi_y)R(z, \phi_z)(\boldsymbol{P}_i - \boldsymbol{C}) \tag{3.2}$$

In the $X''Y''Z''$ coordinate frame, the reference point $\boldsymbol{P}''$ nominally meets (3.3). With (3.2) and (3.4), the optimization problem is defined with (3.5) and k_C $(0 < k_C < 1)$ is a ratio. k_C was set to be 0.5 in this Chapter. The parameters C_x, C_y, C_z, r, ϕ_y and ϕ_z are calculated through solving the optimization problem (3.5) with LMA.

$$\begin{cases} P_{xi}'' = 0.0 \\ \sqrt{{P''_{yi}}^2 + {P''_{zi}}^2} = r \end{cases} \tag{3.3}$$

$$\begin{cases} {}^x f_i = |P_{xi}''| \\ {}^r f_i = |\sqrt{{P''_{yi}}^2 + {P''_{zi}}^2} - r| \end{cases} \tag{3.4}$$

$$\min \sum_{i=1}^{N} ((1 - k_C){}^x f_i^2 + k_C {}^r f_i^2) \tag{3.5}$$

3.1.2 Proposed method: Directly fitting a circular arc in 3D space with joint angle constraint

In this Chapter, it is proposed that a circular arc is directly fitted in 3D space with joint angle constraint. $\tilde{\theta}_u(t)$ denotes the joint angle measured by sensor, where u is the number of the joint which is rotated in the single-joint rotation and t is the number of the reference point position in the single-joint rotation. φ_0 is defined as the artificial initial angle of the circular arc. In the $X''Y''Z''$ coordinate frame, with the joint angle constraint, point $\boldsymbol{P}''$ nominally meets (3.6). So with (3.2) and (3.7), the optimization problem can

be defined with (3.8) and k_D ($0 < k_D < 1$) is a ratio. k_D was set to be 0.5 in this Chapter. The parameters C_x, C_y, C_z, r, ϕ_y, ϕ_z and φ_0 are calculated through solving the optimization problem (3.8) using LMA.

$$\begin{cases} P''_{xi} = 0.0 \\ P''_{yi} = r\cos(\tilde{\theta}_u(t) + \varphi_0) \\ P''_{zi} = r\sin(\tilde{\theta}_u(t) + \varphi_0) \end{cases} \tag{3.6}$$

$$\begin{cases} {}^{x}g_i = |P''_{xi}| \\ {}^{y}g_i = |P''_{yi} - r\cos(\tilde{\theta}_u(t) + \varphi_0)| \\ {}^{z}g_i = |P''_{zi} - r\sin(\tilde{\theta}_u(t) + \varphi_0)| \end{cases} \tag{3.7}$$

$$\min \sum_{i=1}^{N} ((1 - k_D){}^{x}g_i{}^2 + k_D({}^{y}g_i{}^2 + {}^{z}g_i{}^2)) \tag{3.8}$$

The point-plane distance error DE_i is set to be the absolute distance from the point to the estimated plane and calculated with (3.9).

$$DE_i = |\boldsymbol{n}^T \boldsymbol{P}_i + d| \tag{3.9}$$

The function $atan2(y, x)$ is used to calculate the arc tangent of y/x using the signs of arguments to determine the correct quadrant. After directly fitting the circular in 3D space, for the point $\boldsymbol{P}_i$, its radius error RE_i is set to be RE''_i calculated with (3.10) and its angle error AE_i is set to be AE''_i computed with (3.11).

$$RE''_i = |\sqrt{P''_{yi}{}^2 + P''_{zi}{}^2} - r| \tag{3.10}$$

$$AE''_i = |(atan2(P''_{zi}, P''_{yi}) - \tilde{\theta}_u(t)) - (atan2(P''_{z1}, P''_{y1}) - \tilde{\theta}_u(1))| \tag{3.11}$$

3.1.3 Using RANSAC directly fitting a circular arc in 3D space

When the available data points are noisy and rich of outliers which cannot be described by the associated parameters, RANSAC is one of the most

commonly used approaches to fit the data [25].

- A. Using RANSAC directly fitting a circular arc in 3D space with traditional method

  When fitting a circular arc with RANSAC, the traditional method of directly fitting a circular arc in 3D space, which is described in Section 3.1.1, is used for the parameters estimation, and the point-plane distance error DE_i and the radius error RE_i'' are used for the error estimation.

- B. Using RANSAC directly fitting a circular arc in 3D space with joint angle constraint

  When fitting a circular arc with RANSAC, the proposed method of directly fitting a circular arc in 3D space with joint angle constraint, which is described in Section 3.1.2, is used for the parameters estimation, and the point-plane distance error DE_i, the radius error RE_i'' and the joint angle error AE_i'' are used for the error estimation.

3.2 Simulation results

To verify the methods described in this Chapter, the circular arc is fitted in 3D space with the following methods.

- *Method A.* Traditional method of fitting a circular arc through estimating its parameters in 2D space.

  With LMA, the optimal 3D plane can be estimated through solving the optimization problem defined with (2.1) and the normal vector $\boldsymbol{n}$ is obtained. As shown in Figure 2.1, the XYZ coordinate frame is rotated into a new coordinate frame such that its $Y'Z'$ plane is parallel to the estimated plane and $\boldsymbol{n}$ becomes the X' axis.

  In the $X'Y'Z'$ coordinate frame, the position of point $\boldsymbol{P}_i$ is defined as $\boldsymbol{P}'_i(\boldsymbol{P}'_i = [P'_{xi} \quad P'_{yi} \quad P'_{zi}]^T)$ and the center of the circular arc is set to be $\boldsymbol{C}'(\boldsymbol{C}' = [C'_x \quad C'_y \quad C'_z]^T)$. The parameters C'_y, C'_z and r can be

calculated by solving the optimization problem (2.6) using LMA. C'_x is solved with (2.7). After $\boldsymbol{C}'$ is obtained, the center $\boldsymbol{C}$ can be calculated through the rigid transform matrix between the XYZ coordinate frame and the $X'Y'Z'$ coordinate frame.

- *Method B.* Fitting a circular arc through estimating its parameters in 2D space with joint angle constraint.

  With the same method used in the *Method A*, the the normal vector $\boldsymbol{n}$ is calculated. In the $X'Y'Z'$ coordinate frame, with the joint angle constraint, the parameters C'_y, C'_z, r and φ_0 are calculated by solving the optimization problem (2.11) using LMA. C'_x is solved with (2.7). After $\boldsymbol{C}'$ is obtained, the center $\boldsymbol{C}$ can be calculated through the rigid transform matrix between the XYZ coordinate frame and the $X'Y'Z'$ coordinate frame.

- *Method C.* Traditional method of directly fitting a circular arc in 3D space.

- *Method D.* Proposed method: Directly fitting a circular arc in 3D space with joint angle constraint.

- *Method C(R).* Using RANSAC directly fitting a circular arc in 3D space with traditional method.

- *Method D(R).* Using RANSAC directly fitting a circular arc in 3D space with joint angle constraint.

For *Method A* and *Method B*, after fitting a circular arc in 2D space, for the point $\boldsymbol{P}_i$, its radius error RE_i is set to be RE'_i calculated with (3.12) and its angle error AE_i is set to be AE'_i computed with (3.13).

$$RE'_i = |\sqrt{(P'_{yi} - C'_y)^2 + (P'_{zi} - C'_z)^2} - r| \tag{3.12}$$

$$\begin{aligned} AE'_i = |(atan2((P'_{zi} - C'_z), (P'_{yi} - C'_y)) - \tilde{\theta}_u(t)) - \\ (atan2((P'_{z1} - C'_z), (P'_{y1} - C'_y)) - \tilde{\theta}_u(1))| \end{aligned} \tag{3.13}$$

The Root Mean Square (RMS) of point-plane distance error D_S is defined as (2.24) with DE_i. The RMS of radius error R_S is defined as (2.25) with

RE_i and the RMS of joint angle error A_S is defined as (2.26) with AE_i.

Simulations were done to evaluate the performance of the proposed method through comparing it with the previous methods. Besides, using RANSAC the circular arc was fitted with 3D noisy data points. All of the simulations were done on the Matlab platform.

In the simulation, the parameters of the circular arc were set to be: radius $r = 2000.0\ mm$, normal vector $\boldsymbol{n} = [0.9077 \quad -0.2432 \quad 0.3420]^T$, center $\boldsymbol{C} = [-200.0\ mm \quad 300.0\ mm \quad 500.0\ mm]^T$. It is noted the results do not depend on the normal vector or the center or the radius of the circular arc, or the location of the circular arc where the data points are sampled. ΔC and Δr respectively denote the absolute Euclidean distance of the centers and the absolute radius difference between the ground truth and fitting result. $\Delta\theta$ denotes the absolute difference of the angle between the normal vector of the ground truth and fitting result. t is the computation time in milliseconds.

3.2.1 Evaluating the methods of fitting circular arc in 3D space

For the data points, Gaussian noises with zero mean and standard deviation of σ_P were added to the point positions, and Gaussian noises with zero mean and standard deviation of σ_A were added to the joint angle. The circular arcs were fitted with *Method A* to *Method D*. Simulations with the circular arcs of different arc lengths, position noises and angle noises were performed. Only the following results are presented.

1. The degree measure of the circular arc was set to be 45.000 *deg* with the step angle of 0.500 *deg* and 91 data points, and σ_A was set to be 0.020 *deg*. σ_P was changed from 0.5 *mm* to 10.0 *mm*. One example of the data set is presented in Figure 3.2. For the circular arc with each standard deviation σ_P, 1000 circular arcs were simulated and the averaged results from these 1000 data sets were calculated. The simulation results are shown in Figure 3.3.

2. As presented in Figure 3.4, the degree measure of the circular arc was changed from 20.000 *deg* to 80.000 *deg*. For the circular arc of each degree measure, 1000 circular arcs with the step angle of 0.500 *deg* were simulated

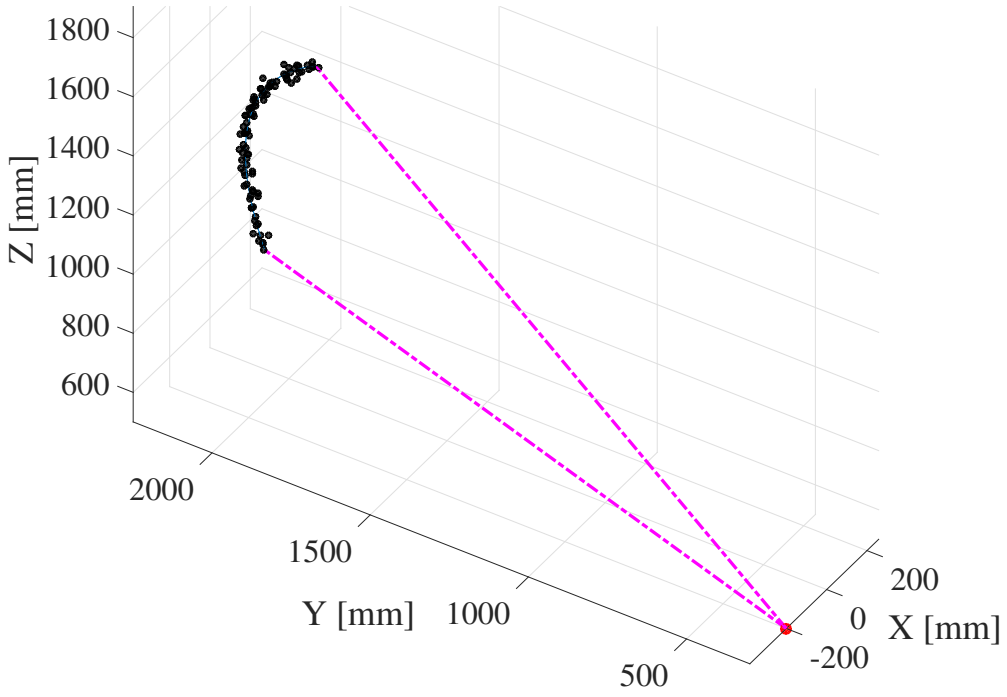
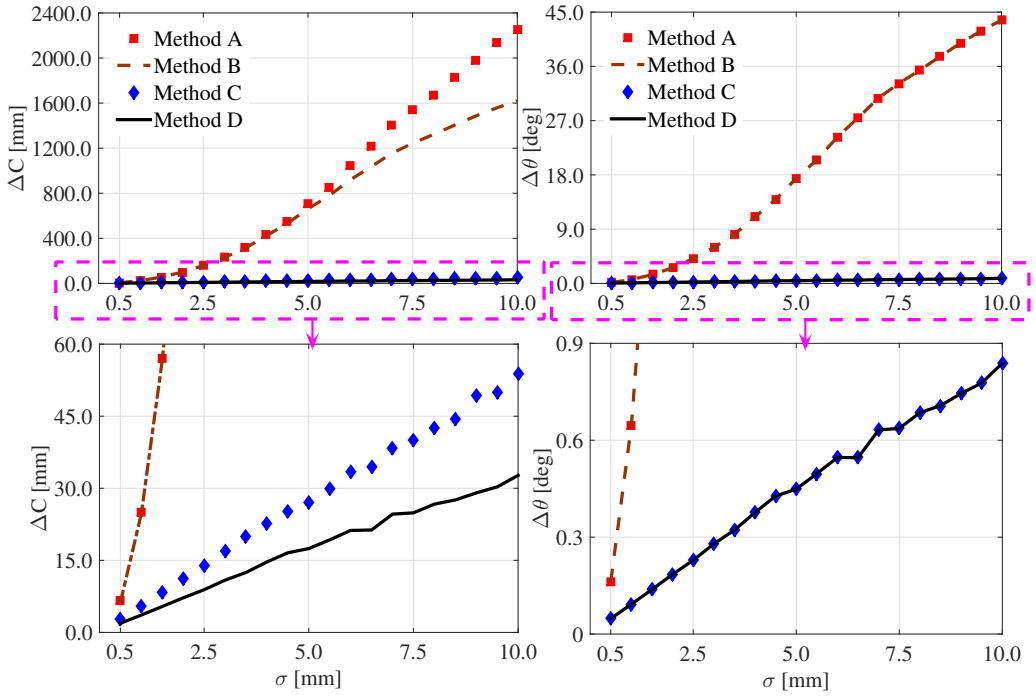


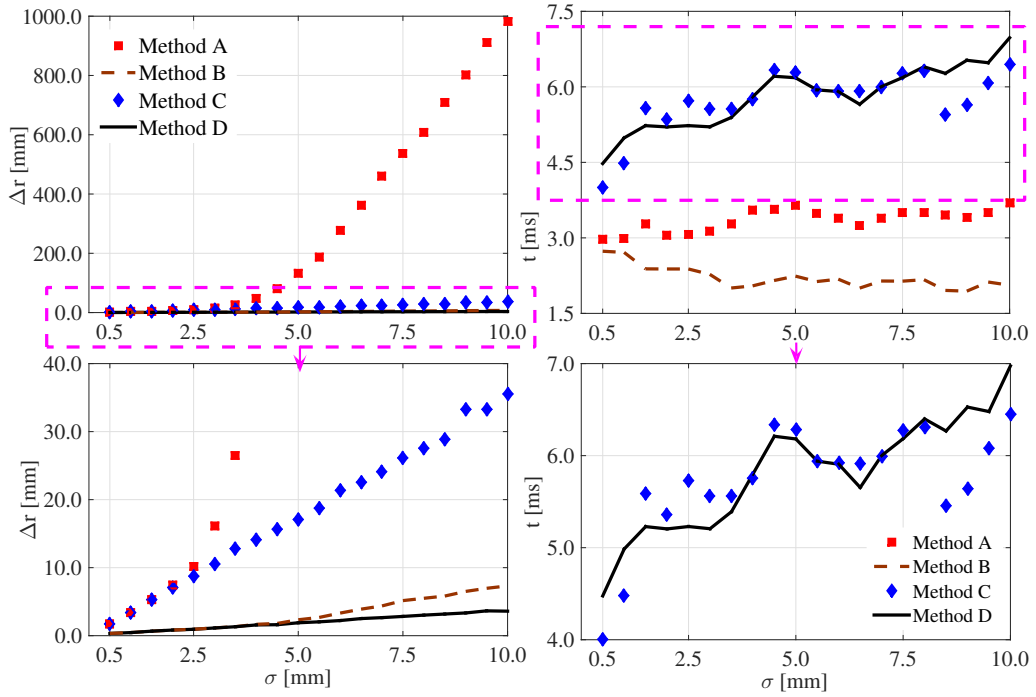
Figure 3.2: A circular arc with the degree measure of 45.000 *deg* with the step angle of 0.500 *deg* and 91 data points.

and the averaged results from these 1000 data sets were calculated. σ_P and σ_A were set to be 6.0 *mm* and 0.020 *deg* respectively. The simulation results are shown in Figure 3.5.

As shown in Figure 3.3 and Figure 3.5, compared with other methods, *Method D* consistently demonstrates superior performance, especially when a short circular arc is fitted and/or the standard deviation σ_P becomes large. For the simulation results shown in Figure 3.5, the averaged D_S and R_S of these circular arcs are presented in Figure 3.6 and the averaged A_S is shown in Figure 3.7. When the circular arc is short and/or the standard deviation σ_P is large, the normal vector of the circular arc can still be well estimated with *Method C* and *Method D* through directly fitting the circular arc in 3D space. However, in this condition it becomes difficult to estimate the normal vector with *Method A* and *Method B* even though their D_S are still small and quite close to the results of *Method C* and *Method D* as shown in Figure 3.6. For *Method B*, since the joint angle is used as a constraint, its A_S is small and



(a) ΔC and $\Delta\theta$



(b) Δr and t

Figure 3.3: The degree measure of the circular arc was set to be 45.000 *deg*. For the circular arc with each standard deviation σ_P, 1000 circular arcs were simulated with Gaussian noises added to the data points and the averaged results from these 1000 data sets were calculated.

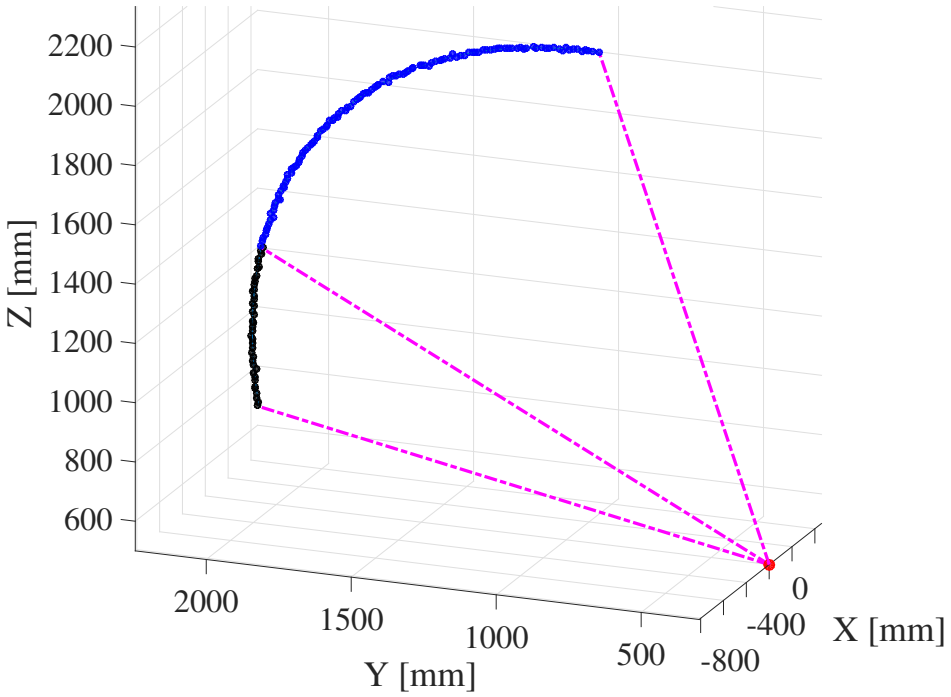


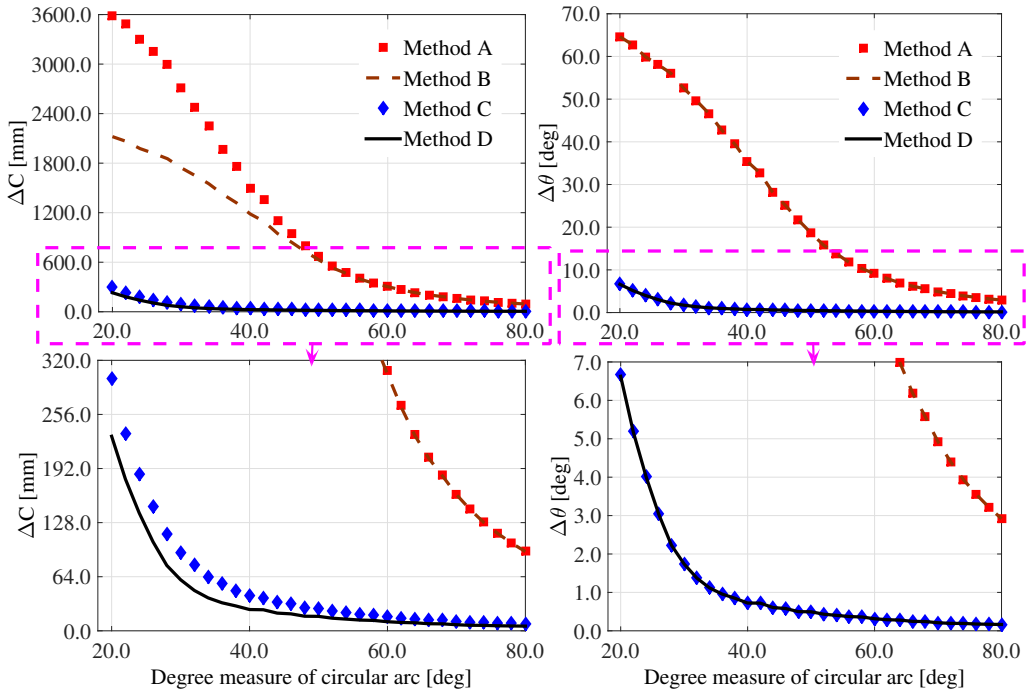
Figure 3.4: The degree measure of the circular arc was changed from 20.000 *deg* to 80.000 *deg* with the step angle of 0.500 *deg*.

its ΔC and Δr become smaller compared to *Method A*. As the joint angle is not used as a constraint in *Method C*, its A_S becomes larger compared to *Method D* and its ΔC and Δr are larger than the results of *Method D* even though their D_S and R_S are quite close.

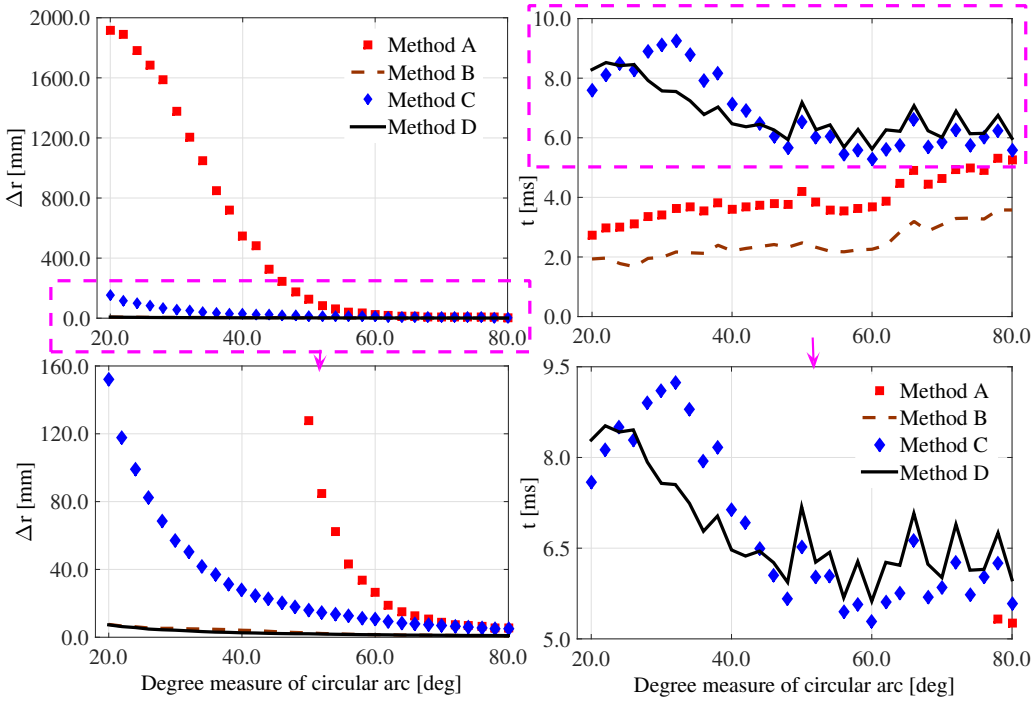
3.2.2 Fitting circular arc in 3D space with RANSAC

The degree measure of the circular arc was set to be 45.000 *deg* with the step angle of 0.500 *deg* and 91 data points, Gaussian noises with zero mean and standard deviation σ_A of 0.020 *deg* were added to the joint angle of the data points, and Gaussian noises with zero mean and standard deviation σ_P of 3.0 *mm* were added to the positions of the data points. For the "outliers", they are randomly located in the space of 800.0 *mm* x 725.0 *mm* x 1375.0 *mm* and N_O denotes the number of the outliers.

Fifty data sets were simulated with the outliers number of N_O and the



(a) ΔC and $\Delta\theta$



(b) Δr and t

Figure 3.5: For the circular arc of each degree measure, 1000 circular arcs were simulated with Gaussian noises added to the data points and the averaged results from these 1000 data sets were calculated.

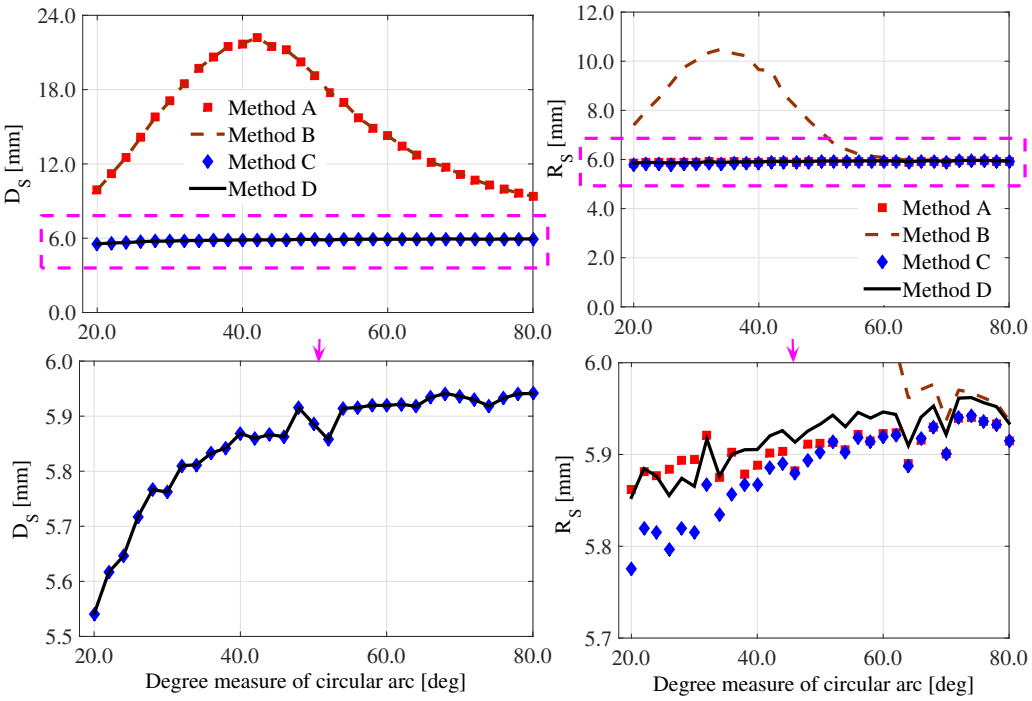


Figure 3.6: The averaged D_S and R_S of the circular arcs fitted in Figure 3.5.

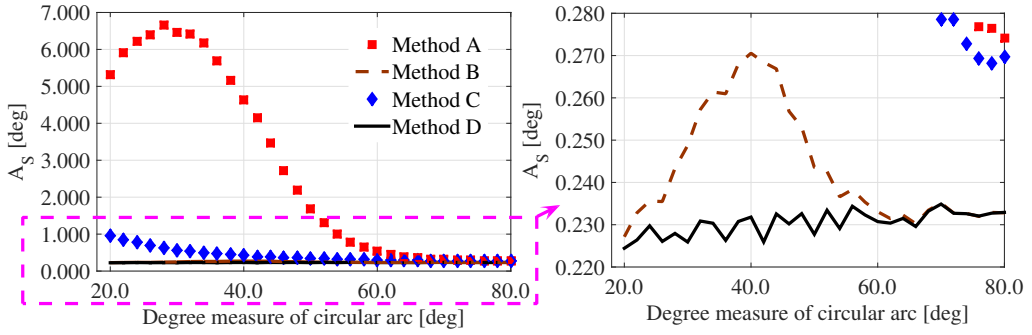


Figure 3.7: The averaged A_S of the circular arcs fitted in Figure 3.5.

averaged results from these 50 data sets were calculated. One example of the data set is shown in Figure 3.8 with N_O of 136. The data points were fitted with *Method C(R)* and *Method D(R)*. The simulation results are presented in Table 3.1 and it shows *Method D(R)* performs better than *Method C(R)* with smaller ΔC, Δr and $\Delta\theta$.

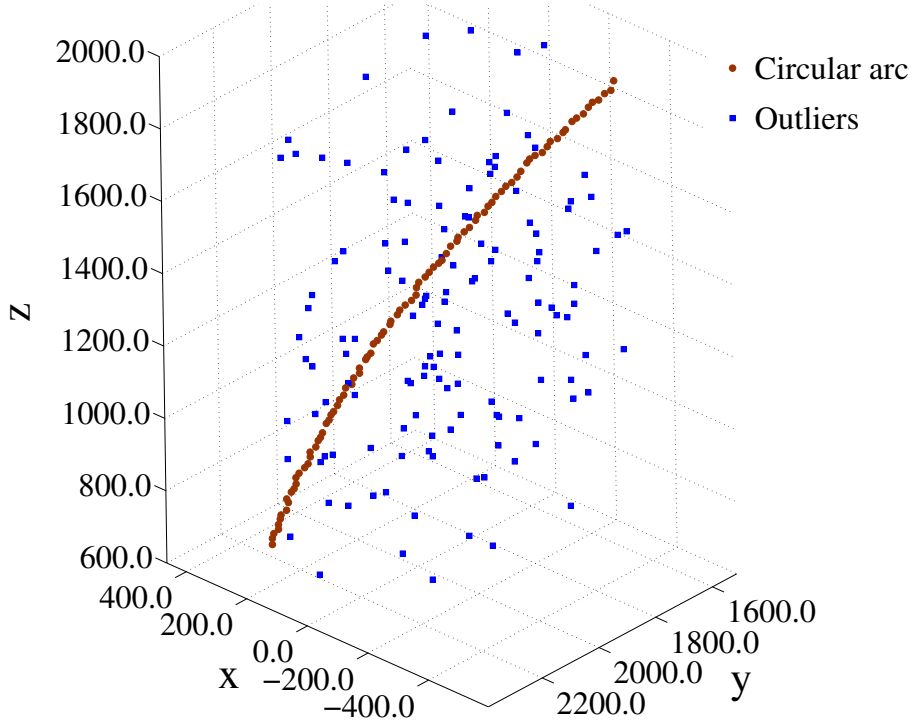


Figure 3.8: A circular arc with 91 data points and 136 outliers.

Table 3.1: Simulations results of fitting circular arc with RANSAC

	Method D(R)			Method C(R)		
N_O	ΔC [mm]	Δr [mm]	$\Delta\theta$ [deg]	ΔC [mm]	Δr [mm]	$\Delta\theta$ [deg]
45	11.088	3.651	0.312	21.472	17.584	0.349
91	12.697	3.645	0.366	26.552	20.794	0.464
136	14.025	4.027	0.380	28.557	24.375	0.395

3.3 Hand-Eye Calibration for the pan-tilt-camera system and experimental results

Whenever a sensor such as a camera is mounted on a robot hand, its position and orientation relative to the hand coordinate frame must be de-

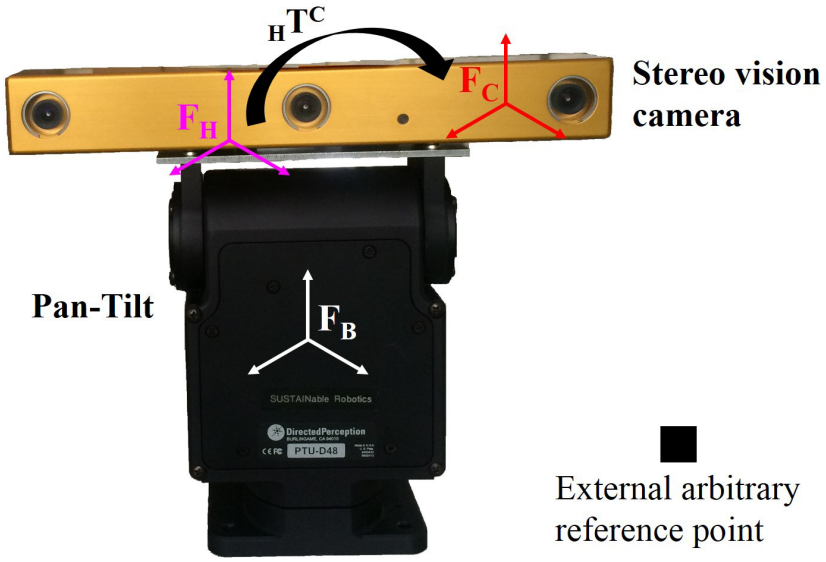


Figure 3.9: The model of the Pan-Tilt-Camera system. The Hand-Eye calibration consists of identifying the unknown relationship between the Hand coordinate frame F_H and the Sensor coordinate frame F_C, which is denoted as ${}_HT^C$ and can be described with the parameters $\boldsymbol{p} = \{\theta_{Cx}, \theta_{Cy}, \theta_{Cz}, P_{Cx}, P_{Cy}, P_{Cz}\}$. In the Hand coordinate frame F_H, θ_{Cx}, θ_{Cy}, θ_{Cz} are the sensor fixing orientation and P_{Cx}, P_{Cy}, P_{Cz} are the sensor fixing position.

termined, and this is normally referred to as the Hand-Eye calibration problem [17, 18]. The model of a Pan-Tilt-Camera system is presented in Figure 3.9. F_B, F_H and F_C are the Pan-Tilt, Hand and Sensor coordinate frame respectively. The Hand-Eye calibration consists of identifying the unknown relationship between the Hand coordinate frame F_H and the Sensor coordinate frame F_C, which is denoted as ${}_HT^C$ and can be described with the parameters $\boldsymbol{p} = \{\theta_{Cx}, \theta_{Cy}, \theta_{Cz}, P_{Cx}, P_{Cy}, P_{Cz}\}$. In the Hand coordinate frame F_H, θ_{Cx}, θ_{Cy}, θ_{Cz} are the sensor fixing orientation and P_{Cx}, P_{Cy}, P_{Cz} are the sensor fixing position.

Pure rotation of the links and circle-point analysis technique were used in robot calibration [7, 13]. In [21], manifolds such as circles were used to do the Hand-Eye calibration using a relatively large number of poses to estimate a

single circle, improving the overall accuracy of calibration. The pure rotation calibration method relies on a sequence of rotations of the links, while using a sensor to track an external arbitrary reference point fixed in Pan-Tilt coordinate frame F_B. In the sensor coordinate frame F_C, the resulting reference point positions measured by sensor are nominally located in a plane whose normal vector is the direction of the rotational axis and describe a circular arc whose center lies on the rotational axis of that joint. The process of identifying the parameters of the Hand-Eye calibration with pure rotation method is divided into two procedures.

Procedure 1. Using the resulting reference point positions measured by the sensor in each single-joint rotation, the normal vector, center and radius of the circular arc are estimated in the sensor coordinate frame F_C.

Procedure 2. Based on the estimated normal vectors, centers and radii of all these single-joint rotations, with their geometry relationship to the calibration parameters $\boldsymbol{p}$, the calibration parameters $\boldsymbol{p}$ are solved [7, 13, 21].

For the pure rotation calibration method, it is critical to estimate the normal vector, center and radius of the circular arc. This Chapter is focused on *Procedure 1.* Traditional methods of fitting the circular arc cannot perform as well as expected when the joint motion that can be measured is limited by the FOV of camera, which is the case studied in this Chapter. Based on the traditional method that uses a sequence of pure rotations, the Hand-Eye calibration is done for the Pan-Tilt-Camera system presented in Figure 3.9. The proposed method of directly fitting the circular arc in 3D space with joint angle constraint is applied in the Hand-Eye calibration process and verified through experiments.

3.3.1 Hand-Eye Calibration using pure rotation method for the Pan-Tilt-Camera system

The pure rotation calibration method is applied to the Pan-Tilt-Camera system with the following steps.

Step 1. Data acquisition.

The position of the external arbitrary reference point is detected by the

sensor and the joint angles are measured through the Pan-Tilt unit. The Tilt angle is changed with $N_1(N_1 \geq 3)$ steps while fixing the Pan angle and the trajectory of the reference point measured by sensor is denoted as motion sequence 1 (Tilt motion). Similarly, the Pan angle is changed with $N_2(N_2 \geq 3)$ steps while fixing the Tilt angle and the trajectory of the reference point measured by sensor is denoted as motion sequence 2 (Pan motion).

Step 2. Parameters calculation.

With the reference point positions of the motion sequence $m(m = 1, 2)$, in the sensor coordinate frame F_C, the normal vector ${}^C\boldsymbol{n}_m^{SC}$ and center ${}^C\boldsymbol{C}_m^{SC}$ of the circular arc are estimated with the circular arc fitting methods described in this Chapter. As the joint axis of the moved joint must pass through the center of this circular arc, from the mechanical structure, the theoretical rotation axis $\boldsymbol{n}_m^T$ and theoretical rotation center $\boldsymbol{C}_m^T$ of the motion sequence m are known. Using their constant geometry relationships to ${}^C\boldsymbol{n}_m^{SC}$ and ${}^C\boldsymbol{n}_m^{SC}$, the calibration parameters $\boldsymbol{p}$ can be solved [7, 13, 21].

3.3.2 Experimental results

The Pan-Tilt-Camera system shown in Figure 3.9 was used to verify the proposed calibration method. The Pan-Tilt unit is a Pan-Tilt Unit D48 (PTU-D48) from FLIR Motion Control Systems. The Pan resolution is 0.006 *deg* and the Tilt resolution is 0.003 *deg*. The sensor is a Bumblebee XB3 stereo vision camera from Point Grey Research and its HFOV is 66.0 *deg*. A chessboard with the size of 3 x 4 was used and its center was considered as the external reference point.

With the *Data acquisition* method described in this Chapter, the Bumblebee XB3 stereo vision camera was used to acquire the positions of the external reference point. The range of Tilt motion is 38.571 *deg* with 14 sampled points and step angle of 2.967 *deg*. The range of Pan motion is 48.240 *deg* with 16 sampled points and step angle of 3.216 *deg*. The circular arcs were fitted with *Method A* to *Method D(R)* and the calibration parameters $\boldsymbol{p}$ were solved with the *Parameters calculation* method described in this Chapter.

After calibration, for the identification points, their positions in the Pan-Tilt coordinate frame are calculated and denoted as ${}^{I}\boldsymbol{P}^{B}_{m,i}$. In the Pan-Tilt coordinate frame, as the external reference point was fixed, its averaged position $\overline{{}^{I}\boldsymbol{P}^{B}_{ext}}$ is estimated with (3.14) and the value E_I defined with (3.15) is in terms of RMS difference between $\overline{{}^{I}\boldsymbol{P}^{B}_{ext}}$ and ${}^{I}\boldsymbol{P}^{B}_{m,i}$.

$$\overline{{}^{I}\boldsymbol{P}^{B}_{ext}} = \frac{1}{\sum_{m=1}^{2} N_m}\left(\sum_{m=1}^{2}\sum_{i=1}^{N_m} {}^{I}\boldsymbol{P}^{B}_{m,i}\right) \tag{3.14}$$

$$E_I = \sqrt{\frac{1}{\sum_{m=1}^{2} N_m}\left(\sum_{m=1}^{2}\sum_{i=1}^{N_m} \left|{}^{I}\boldsymbol{P}^{B}_{m,i} - \overline{{}^{I}\boldsymbol{P}^{B}_{ext}}\right|^2\right)} \tag{3.15}$$

In order to verify the calibration results, the chessboard was placed to M_V (M_V was 6 in this Chapter) different random positions. For each position, in the range that chessboard can be detected by the stereo vision camera, the Pan angle and Tilt angle were randomly set and the stereo vision camera viewed the chessboard with L_k ($k = 1, 2, ..., M_V$) different poses (L_k was 10 in this Chapter). Similarly, in the Pan-Tilt coordinate frame, the positions of the validation points are calculated and denoted as ${}^{V}\boldsymbol{P}^{B}_{k,i}$. In the Pan-Tilt coordinate frame, for each chessboard position, its averaged position $\overline{{}^{V}\boldsymbol{P}^{B}_{k}}$ is estimated with (3.16) and the value E_V defined with (3.17) is in terms of RMS difference between $\overline{{}^{V}\boldsymbol{P}^{B}_{k}}$ and ${}^{V}\boldsymbol{P}^{B}_{k,i}$.

$$\overline{{}^{V}\boldsymbol{P}^{B}_{k}} = \frac{1}{L_k}\left(\sum_{i=1}^{L_k} {}^{V}\boldsymbol{P}^{B}_{k,i}\right) \tag{3.16}$$

$$E_V = \sqrt{\frac{1}{\sum_{k=1}^{M_V} L_k}\left(\sum_{k=1}^{M_V}\sum_{i=1}^{L_k} \left|{}^{V}\boldsymbol{P}^{B}_{k,i} - \overline{{}^{V}\boldsymbol{P}^{B}_{k}}\right|^2\right)} \tag{3.17}$$

The experimental results are shown in Table 3.2. When the circular arcs were fitted with *Method A* to *Method D*, it is noticed that the results of *Method B* and *Method D* are better than the results of *Method A* and *Method C* both for the identification data and for the validation data. Better results are obtained when RANSAC is used in *Method C(R)* and *Method D(R)* to fit

the circular arcs as the "outliers" could be removed. However, *Method D(R)* still performs better than *Method C(R)* both for the identification data and for the validation data. So when the joint angle is added as a constraint, the result is consistently better than the results calculated without joint angle constraint.

Table 3.2: Experimental results

Method	Identification, E_I [mm]	Validation, E_V [mm]
A	2.188	1.975
B	1.408	1.319
C	2.188	1.975
D	1.408	1.320
C(R)	1.366	1.285
D(R)	1.354	1.275

3.4 Conclusion

Through the simulation results and experimental results, it is proved that using the proposed method of directly fitting circular arc in 3D space with joint angle constraint, a circular arc can be fitted more precisely in 3D space compared with the traditional methods, especially when it is a short arc and/or the measured data is noisy. When using the pure rotation calibration method to do the Hand-Eye calibration for a Pan-Tilt-Camera system, the proposed methods perform better than the traditional methods. The proposed method of directly fitting a circular arc in 3D space with joint angle constraint can be used in other applications when the actual joint angle can be measured.

Chapter 4

Terrain Mapping under Extreme Light Conditions with Direct Stereo Matching Method

As shown in Figure 4.2, through multiple exposures, the dynamic range of images could be increased. In [34], a system architecture was introduced for terrain mapping using stereo vision camera. As presented in Figure 4.1, traditionally, the 2D images captured with multiple exposures are fused with Exposure Fusion [36]. With the resulting fused images, the disparity image is computed through stereo matching and the 3D terrain map is reconstructed. In this Chapter, instead of using existing lighting enhancement methods such as Exposure Fusion to increase the texture of 2D image, the stereo matching was directly done using the images grabbed with multiple exposures and the 2D image process of exposure fusion is not needed.

Since it is possible that the camera is moved when the images are grabbed with multiple exposures, the images grabbed with short and long exposures are aligned to the image captured with auto exposure. The matching costs of the resulting registered images and the image grabbed with auto exposure are directly summed by weight. In order to evaluate the performance of

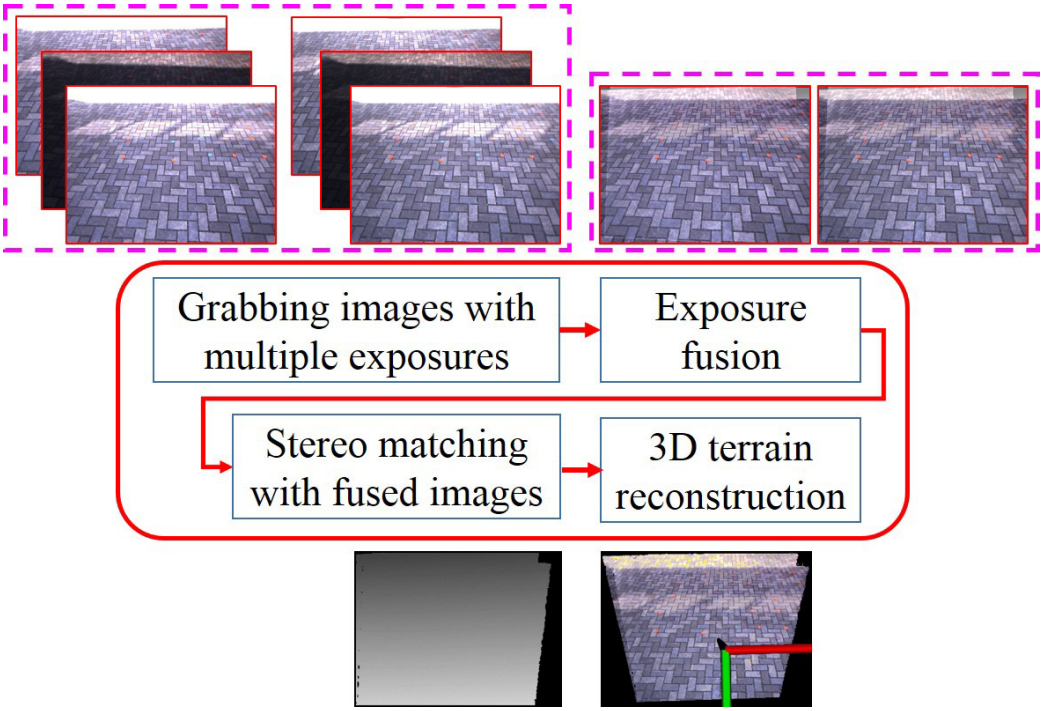


Figure 4.1: The previous system architecture for 3D terrain mapping using a stereo vision camera.

the proposed method, two different stereo matching algorithms were used: a local window-based method and semi-global method. Through experiments in laboratory and outdoors with a stereo vision camera fixed on a tripod and held in the hand, it was verified that with the proposed method, more valid 3D points could be obtained and the terrain maps could be reconstructed more accurately. Especially when the local window-based method was used, the proposed method performed much better than the traditional methods. The remainder of the Chapter is structured as follows. Section 4.1 introduces the proposed methods in detail and the experimental results are presented in Section 4.2. In section 4.3 the conclusion is presented.

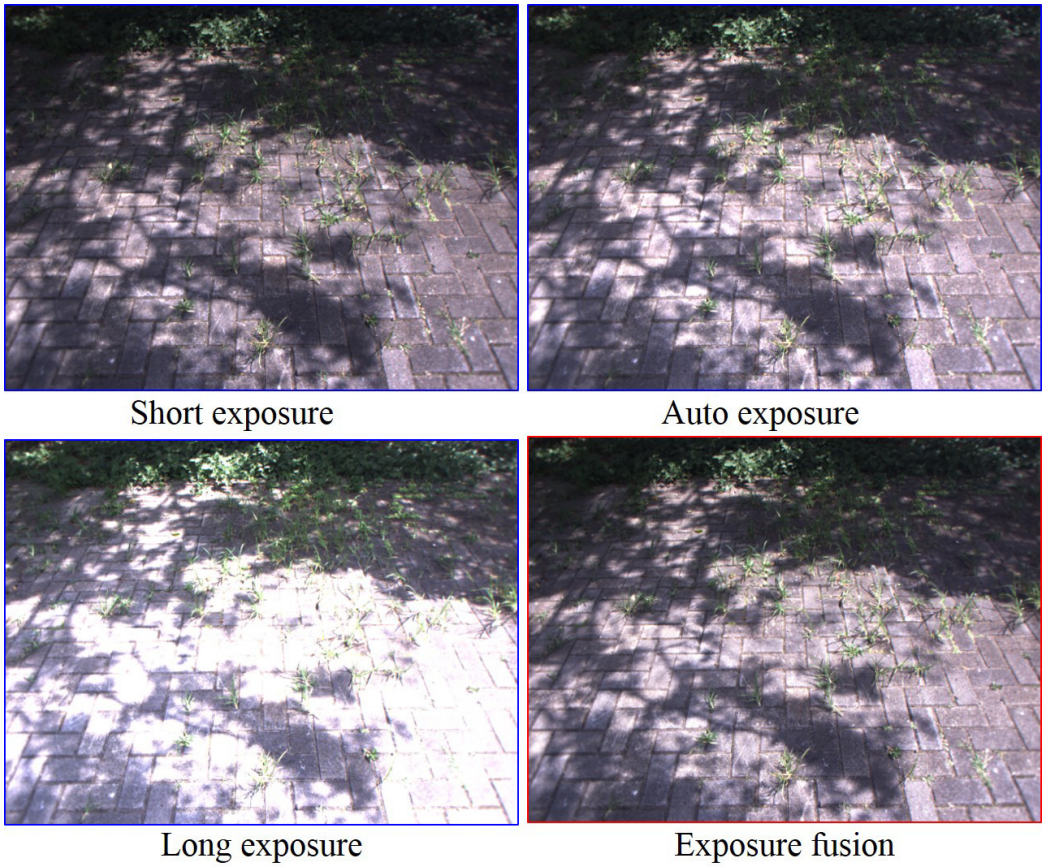


Figure 4.2: The images were grabbed with different shutter times. Auto mode: 0.821 milliseconds. Short exposure: 0.755 milliseconds. Long exposure: 1.806 milliseconds. With exposure fusion, these images were fused.

4.1 Proposed method

The proposed system architecture for 3D terrain mapping with a stereo vision camera is presented in Figure 4.3 and introduced in detail in this section. The images grabbed with multiple exposures are aligned to the image captured with auto exposure. Compared to the previous system architecture presented in Figure 4.1, for the proposed system architecture, stereo matching is directly done with the resulting registered images and Exposure Fusion is not needed.

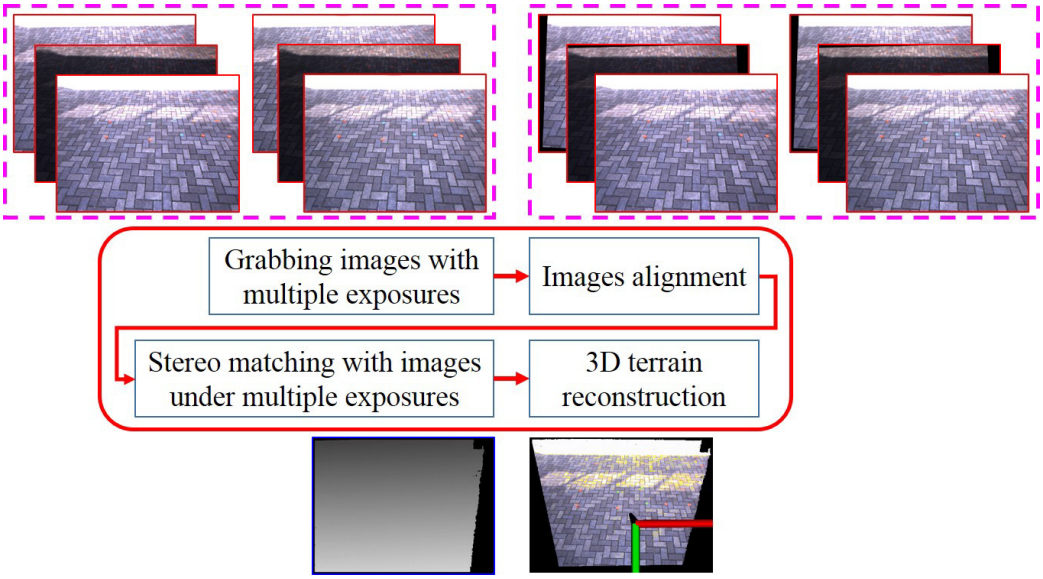


Figure 4.3: The proposed system architecture for 3D terrain mapping using a stereo vision camera.

4.1.1 Acquiring images with multiple exposures

In order to acquire the images with multiple exposures, it is important to properly set the exposure parameters of camera, alternating between a long exposure to capture the shadows and a short exposure to capture the highlights. Using a method similar with [39], the shutter times for short exposure and long exposure are set. T_{min} and T_{max} are the available minimum and maximum shutter times of the camera respectively. Through the API of Bumblebee stereo vision camera from Point Grey Research, the user could get the shutter time T_{auto} in auto exposure, minimum shutter time T_{min} and maximum shutter time T_{max}. Based on the shutter time T_{auto} in auto exposure, the shutter time T_{short} of short exposure and the shutter time T_{long} of long exposure are set with (4.1) and (4.2) respectively. k_{max} is the user defined maximum ratio for k_{long}. For the short exposure, it is required that fewer than p_{short} (e.g. 1%) of the pixels in the image are bright which have values above B_{short} (e.g. 217). If there are too many bright pixels, the exposure time is decreased for the subsequent short exposures. Similarly, for the long exposure it is required that fewer than p_{long} (e.g. 1%) of pixels are

dark which have values less than B_{long} (e.g. 38), otherwise the exposure time is increased for the subsequent long exposures. The camera gain is kept as low as possible to minimize noise, only raising it when the camera shutter time setting is not available for the camera [39]. One example is shown in Figure 4.2, and the images with the short exposure and long exposure were grabbed with this method.

$$T_{short} = k_{short} T_{auto} \tag{4.1}$$

where

$$0 < k_{short} \leq 1, \qquad T_{min} \leq T_{short}$$

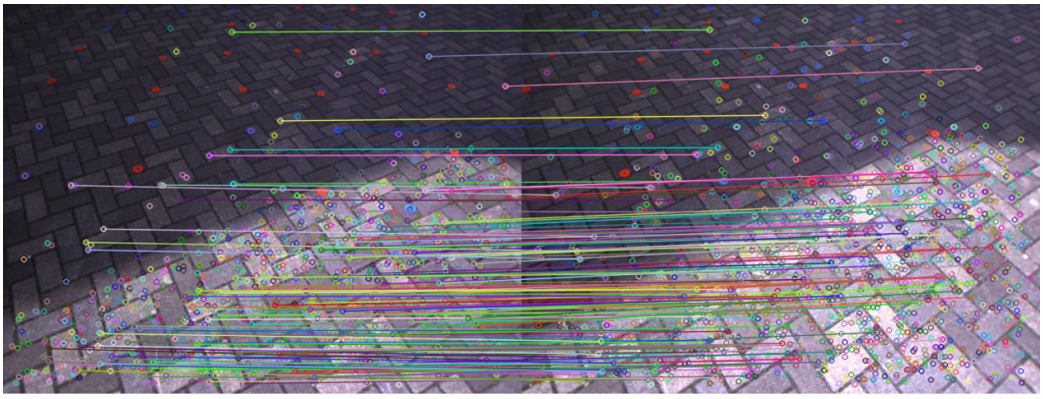
$$T_{long} = k_{long} T_{auto} \tag{4.2}$$

where

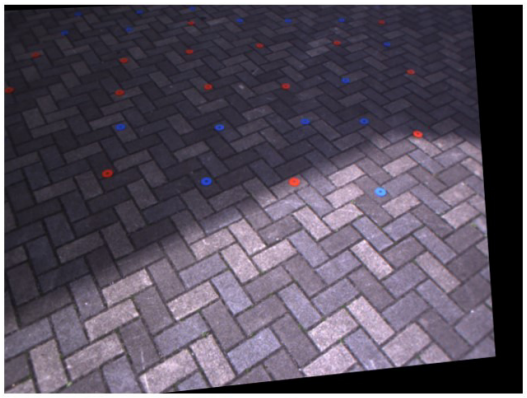
$$1 \leq k_{long} \leq k_{max}, \qquad T_{long} \leq T_{max}$$

4.1.2 Image alignment

In the real application, since it is possible that the camera is moved between exposures when the images are grabbed, it is important to register the images. With a method similar to [42, 43], the correspondence relationship between the images grabbed with short (or long) exposure and auto exposure are calculated with image alignment algorithm and the image acquired with short (or long) exposure is aligned to the image acquired with auto exposure. As shown in Figure 4.4, the key points of the images grabbed in short (or long) exposure and auto exposure are detected with Speeded Up Robust Features (SURF) [29]. After matching the descriptor vectors of the key points, the key points pairs are obtained and denoted as (x_i, y_i) in the image with short (or long) exposure and $(x_i^{'}, y_i)^{'}$ in the image with auto exposure. They are related with (4.3), where $\boldsymbol{H}$ is an arbitrary 3x3 matrix and itself homogeneous [41]. RANSAC is performed to estimate the homography matrix $\boldsymbol{H}$ through solving the optimization problem (4.4), where N_R is the number of the key points pairs which are used to estimate the parameter $\boldsymbol{H}$. The image



Key points detected with SURF and matched pairs of key points



Aligned image

Figure 4.4: The key points of the images grabbed in short (or long) exposure and auto exposure are detected with Speeded Up Robust Features (SURF). After matching the descriptor vectors of the key points, the key points pairs are obtained. RANSAC is performed to estimate the homography matrix and the image acquired with short (or long) exposure is aligned to the image acquired with auto exposure according to the homography matrix.

acquired with short (or long) exposure is aligned to the image acquired with auto exposure according to the homography matrix $\boldsymbol{H}$.

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \sim \boldsymbol{H} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{4.3}$$

with

$$\boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

$$\min \sum_{i=1}^{N_R} (x_i^{'} - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}})^2 + (y_i^{'} - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}})^2 \quad (4.4)$$

4.1.3 Proposed method of matching cost aggregation: Summing the matching costs of the images grabbed with multiple exposures by weight

In this Chapter, the matching cost was defined based on intensity (luminance) instead of colour, which were stored as 8-bit unsigned integers. It is simple to extend this matching cost to colour by computing the costs for each colour channel separately and then summing the matching costs over all channels. The matching cost can be calculated with the methods such as Absolute Difference (AD), Squared Difference (SD), Census Transform (CT) and so on. For the image of the $k-th$ image (auto, short and long exposures in sequence, $k = 1, 2, 3$), for each pixel $\boldsymbol{p}$ ($\boldsymbol{p} = (x, y)$), its matching cost is defined to be $V_{\boldsymbol{p},k}$. As shown in Figure 4.5, the matching costs $V_{\boldsymbol{p},1}$, $V_{\boldsymbol{p},2}$ and $V_{\boldsymbol{p},3}$ for the images captured with auto, short and long exposures are calculated respectively.

For the grayscale image of the $k - th$ image, the intensity of the pixel $\boldsymbol{p}$ is defined as $I(\boldsymbol{p}, k)$ $(0 \leq I(\boldsymbol{p}, k) \leq 255)$ and the exposure quality $\phi^e_{\boldsymbol{p},k}$ is calculated based on how close it is to 127.5 with (4.5) [36]. σ_e was set to be 0.2 in this Chapter. The exposure quality weight $w^e_{\boldsymbol{p},k}$ is computed with (4.6).

$$\phi^e_{\boldsymbol{p},k} = exp(-\frac{(I(\boldsymbol{p}, k) - 127.5)^2}{2(255\sigma_e)^2}) \quad (4.5)$$

$$w^e_{\boldsymbol{p},k} = \frac{\phi^e_{\boldsymbol{p},k}}{\sum_{k=1}^{3} \phi^e_{\boldsymbol{p},k}} \quad (4.6)$$

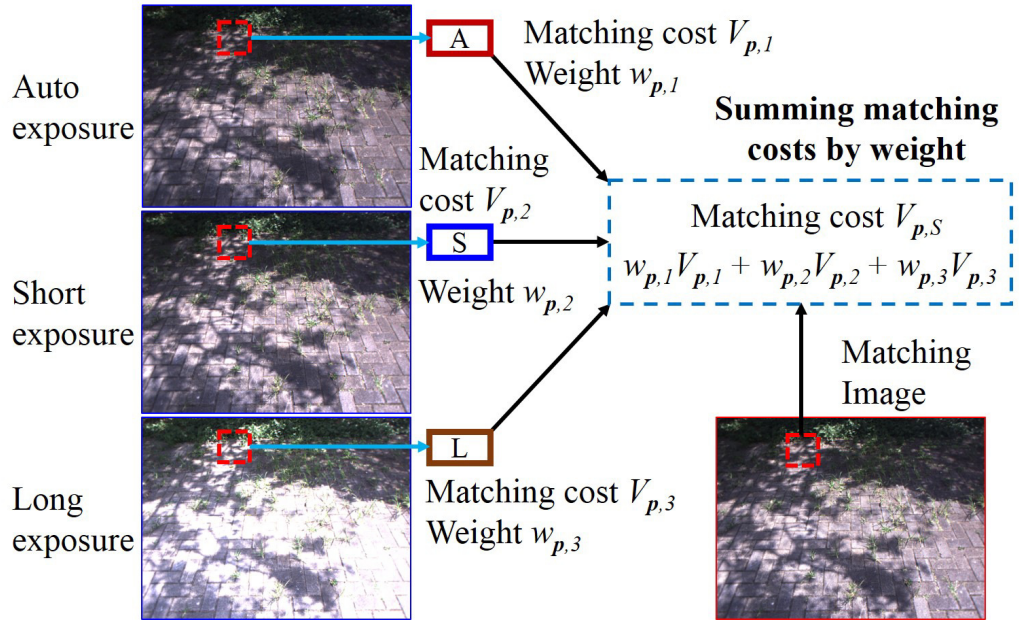


Figure 4.5: For each pixel $\boldsymbol{p}$, the matching costs $V_{\boldsymbol{p},1}$, $V_{\boldsymbol{p},2}$ and $V_{\boldsymbol{p},3}$ for the images captured with auto, short and long exposures can be calculated respectively with the methods such as Absolute Difference (AD), Squared Difference (SD), Census Transform (CT) and so on. For the matching image, the matching cost $V_{\boldsymbol{p},S}$ is directly summed by weight.

For the pixel $\boldsymbol{p}$, $N(\boldsymbol{p})$ is the set of pixels surrounding it in the neighborhood with the window size of L_w x L_h pixels, where L_w and L_h are the width and height of the window in pixels respectively. For the grayscale image of the $k-th$ image, the intensity difference of the pixel $\boldsymbol{p}$ is defined as $S^c_{\boldsymbol{p},k}$ and calculated with (4.7) by comparing its intensity with the pixels in its local neighborhood $N(\boldsymbol{p})$. The intensity diversity $\phi^c_{\boldsymbol{p},k}$ is calculated with (4.8) and σ_c was set to be 0.2 in this Chapter. The intensity diversity weight $w^c_{\boldsymbol{p},k}$ is computed with (4.9).

$$S^c_{\boldsymbol{p},k} = \sum_{\boldsymbol{q}\in N(\boldsymbol{p})} f(I(\boldsymbol{p},k), I(\boldsymbol{q},k)), \tag{4.7}$$

where

$$f(x,y) = \begin{cases} 1 & \text{if } x < y, \\ 0 & \text{else.} \end{cases}$$

$$\phi^c_{\boldsymbol{p},k} \;=\; exp(-\frac{(S^c_{\boldsymbol{p},k} - 0.5L_c)^2}{2(L_c\sigma_c)^2}), \tag{4.8}$$

where

$$L_c = L_w L_h - 1.$$

$$w^c_{\boldsymbol{p},k} \;=\; \frac{\phi^c_{\boldsymbol{p},k}}{\sum_{k=1}^{3}\phi^c_{\boldsymbol{p},k}} \tag{4.9}$$

For the pixel $\boldsymbol{p}$ in the $k-th$ image, its weight $w_{\boldsymbol{p},k}$ is calculated with (4.10) through summing the exposure quality weight $w^e_{\boldsymbol{p},k}$ and intensity diversity weight $w^c_{\boldsymbol{p},k}$, and λ_c was set to be 0.1 in this Chapter. As shown in Figure 4.5, for each pixel $\boldsymbol{p}$, the matching cost of the matching image is defined as $V_{\boldsymbol{p},S}$ and calculated with (4.11) based on the matching costs $V_{\boldsymbol{p},1}$, $V_{\boldsymbol{p},2}$, $V_{\boldsymbol{p},3}$ and the weights $w_{\boldsymbol{p},1}$, $w_{\boldsymbol{p},2}$, $w_{\boldsymbol{p},3}$.

$$w_{\boldsymbol{p},k} \;=\; w^e_{\boldsymbol{p},k} + \lambda_c w^c_{\boldsymbol{p},k} \tag{4.10}$$

$$V_{\boldsymbol{p},S} = \sum_{k=1}^{3} w_{\boldsymbol{p},k} V_{\boldsymbol{p},k} \tag{4.11}$$

One example is shown in Figure 4.6. For each pixel $\boldsymbol{p}$ of the images grabbed with auto, short and long exposures, its weight $w_{\boldsymbol{p},k}$ was calculated with the proposed method respectively. The intensity diversity weight of each pixel was calculated with a window of 15x15 pixels. In the weight image, for each pixel $\boldsymbol{p}$, the red colour means that the auto exposure image has the biggest weight, the green colour means that the short exposure image has the biggest weight and the blue colour means that the long exposure image has the biggest weight.

4.2 Experimental results

In order to evaluate the performance of the stereo matching algorithm described in this Chapter, a Bumblebee XB3 stereo vision camera from Point

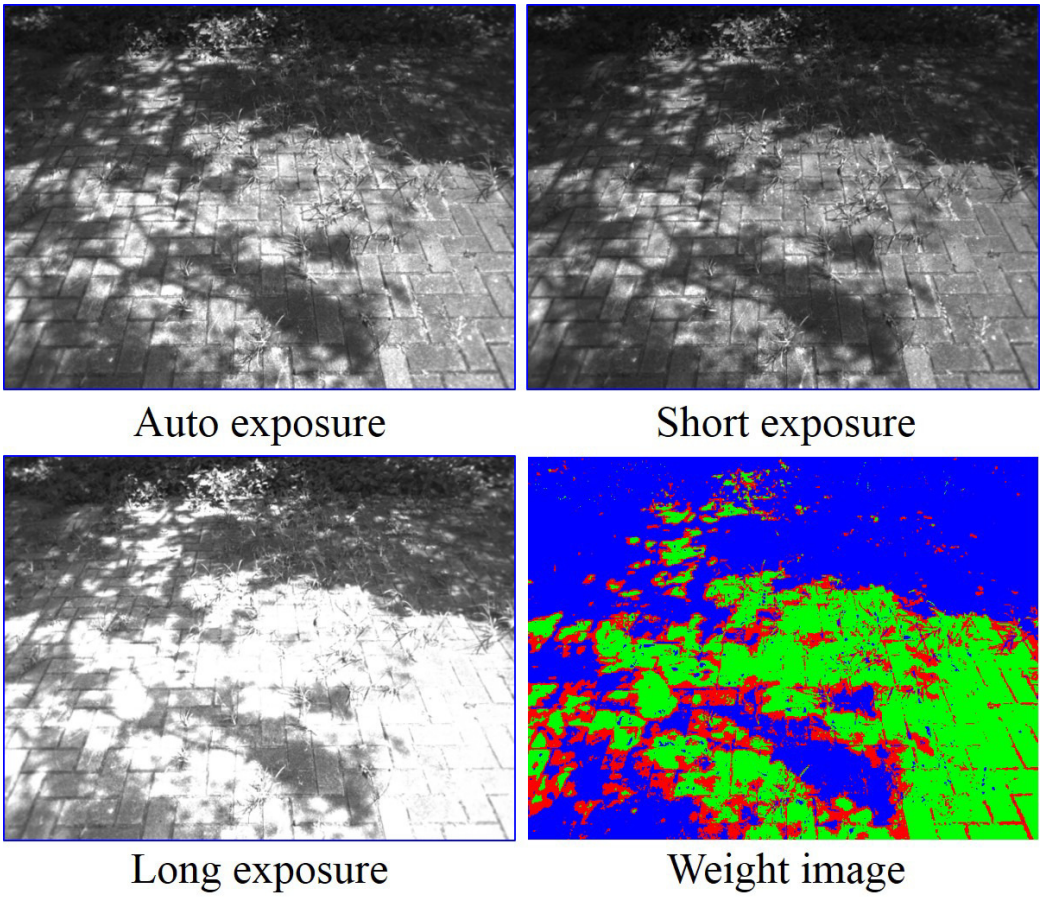


Figure 4.6: For each pixel $\boldsymbol{p}$ of the images grabbed with auto, short and long exposures, its weight $w_{\boldsymbol{p},k}$ was calculated with the proposed method respectively. In the weight image, for each pixel $\boldsymbol{p}$, the red colour means that the auto exposure image has the biggest weight, the green colour means that the short exposure image has the biggest weight and the blue colour means that the long exposure image has the biggest weight. The intensity diversity weight of each pixel was calculated with a window of 15x15 pixels.

Grey Research as shown in Figure 4.7, was used in the experiments and the experiments were done in laboratory and outdoors. The Bumblebee XB3 stereo vision camera is a 3-sensor multi-baseline IEEE-1394b (800Mb/s) stereo vision camera designed for improved accuracy and pre-calibrated for lens distortions and camera misalignments, The colour images grabbed with stereo vision camera were converted to gray images, and the matching costs

Figure 4.7: The Bumblebee XB3 stereo vision camera.

of the matching images were calculated with the gray images. The image size is 640x480 pixels. The stereo matching with a disparity range of 100 pixels was done with local window-based method and SGM respectively.

4.2.1 Stereo matching methods

Census transform was used to calculate the matching cost in this Chapter. It is able to deal with radiometric changes since it is a non-parametric local transform which relies on the relative ordering of local intensity values and not on the intensity values themselves [56]. The census transform encodes the local neighborhood (e.g. window with a window size of 11x11 pixels) around each pixel into a bit cost that only stores whether the compared neighboring pixel has a lower value than the center pixel or not. For the matching image (left image or right image of a stereo pair), its matching cost between two pixels in the matching image and reference image of a stereo pair is the Hamming distance of their census transform in their local windows.

For each pixel $\boldsymbol{p}$, the matching cost of the matching image is defined as $V_{\boldsymbol{p},M}$. For the image grabbed with auto exposure, for each pixel $\boldsymbol{p}$, its matching cost $V_{\boldsymbol{p},1}$ was computed. Using the image alignment method described in this Chapter, the images grabbed with short and long exposures were aligned to the image grabbed with auto exposure. For the registered images of the photographs captured with short and long exposures, its matching

costs $V_{\boldsymbol{p},2}$ and $V_{\boldsymbol{p},3}$ were computed. The image grabbed with auto exposure and the registered images of the photographs captured with short and long exposures were fused with exposure fusion [36] for the resulting fused image and the matching cost of each pixel $\boldsymbol{p}$ is defined as $V_{\boldsymbol{p},E}$. In order to evaluate the performance of the proposed method, for each pixel $\boldsymbol{p}$, the matching cost $V_{\boldsymbol{p},M}$ is calculated with the following four methods.

- Auto exposure. $V_{\boldsymbol{p},M}$ is set to be $V_{\boldsymbol{p},1}$.
- Exposure fusion. $V_{\boldsymbol{p},M}$ is set to be $V_{\boldsymbol{p},E}$.
- Multiple images. $V_{\boldsymbol{p},M}$ is set to be $V_{\boldsymbol{p},D}$, which is calculated with (4.12) through directly summing the matching costs $V_{\boldsymbol{p},1}$, $V_{\boldsymbol{p},2}$ and $V_{\boldsymbol{p},3}$. .

$$V_{\boldsymbol{p},D} = \sum_{k=1}^{3} V_{\boldsymbol{p},k} \tag{4.12}$$

- Multiple images by weight. $V_{\boldsymbol{p},M}$ is set to be $V_{\boldsymbol{p},S}$, which is calculated with (4.11) using the proposed method. The intensity diversity weight of each pixel was calculated with same window size of census transform.

As the performance of a matching cost depends on the algorithm which uses it, two different stereo algorithms were used: a local window-based method [56] and semi-global matching (SGM) [57]. For the local window-based method, after computing the matching costs, the disparity with the lowest summed cost was selected with winner-takes-all. The SGM is adopted as the optimization technique to stereo matching for it is more advantageous since it delivers denser results with far fewer outliers. Many applications have proved that SGM is of high quality and can reconstruct thin or small objects. In this Chapter, the stereo matching was done with SGM by summing the matching costs in four directions (up, down, left and right).

With the method described in [58], the sub-pixel disparity refinement is obtained through interpolating the three matching costs (the winning cost value and its neighbors). The occlusions and mismatches are distinguished by the left/right consistency check, which invalidates disparities if the disparity

with the left/right images stereo matching and its corresponding disparity of with the right/left images stereo matching differ by more than 1 pixel. The uniqueness check invalidates disparities if the minimum cost is not unique.

4.2.2 Experiments of mapping a flat terrain

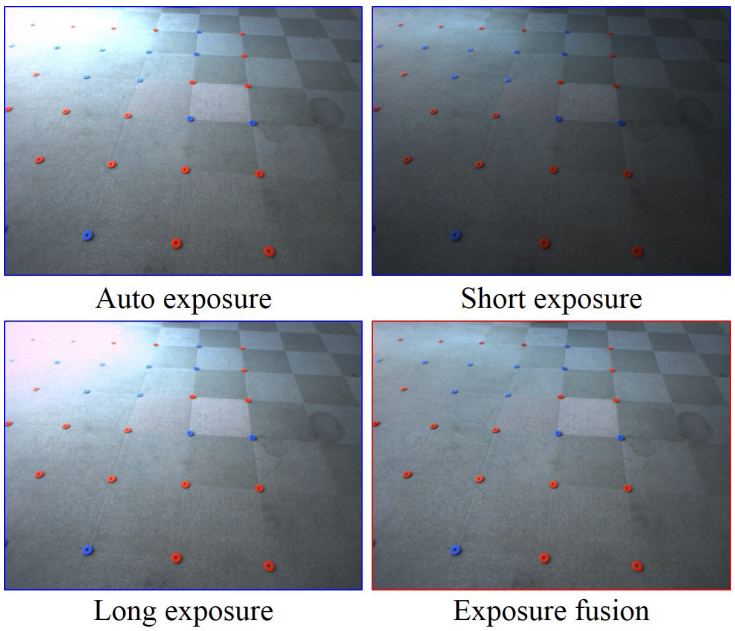
Since the terrain for mapping is almost flat, in order to evaluate the stereo matching result, the best-fit plane was estimated with RANSAC. The distance of the point to the estimated plane is defined as d_t. Points with distance d_t smaller than D_T were considered to be valid. D_T was set to be 40.0 mm in this Chapter. The number of valid points is defined to be N_V, which is the most important criteria to evaluate the performance of the stereo matching methods. The average distance to the estimated plane is define as d_V and calculated with (4.13) using the valid points.

$$d_V = \sqrt{\frac{1}{N_V} \sum_{|d_t|<D_T} {d_t}^2} \tag{4.13}$$

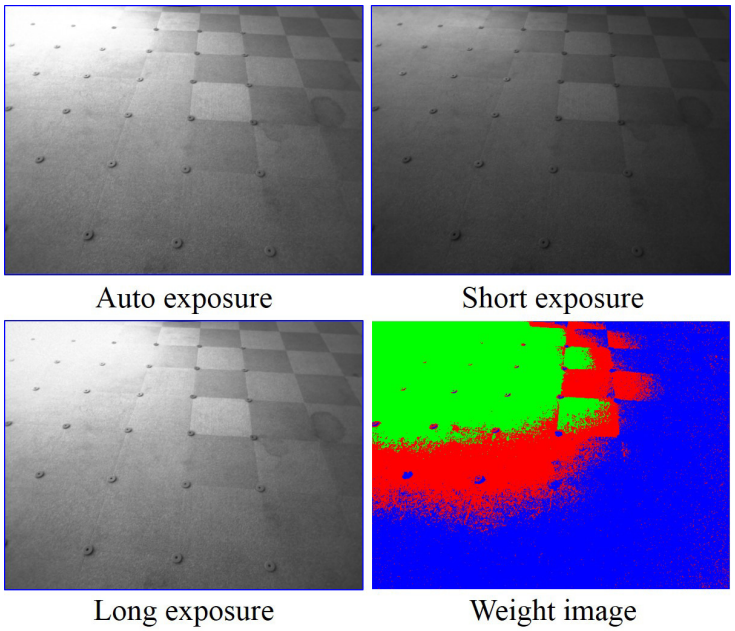
A. Experiments in laboratory with the stereo vision camera fixed on a tripod

The experiments were done in laboratory with the stereo vision camera fixed on a tripod and the grabbed images of the right camera are presented in Figure 4.8(a). The image grabbed with auto, short and long exposures were fused with exposure fusion and the resulting fused image is shown in Figure 4.8(a). The colour images were converted to grayscale images as shown in Figure 4.8(b). As an example, using the method described in this Chapter, the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels and the weight image is shown in Figure 4.8(b).

First, stereo matching was done with the local window-based method. For example, with a window of 15x15 pixels, the disparity images calculated with four different methods are shown in Figure 4.9(a). As it is overexposure and texture-less in the top left of the image grabbed with auto exposure, with the method "auto exposure", the disparity values for the pixels in this

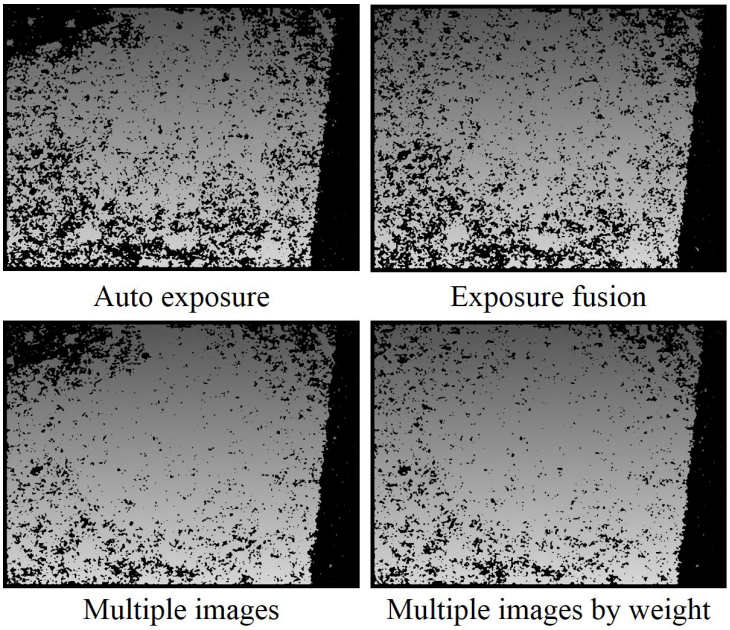


(a) The images were grabbed with different shutter times. Auto exposure: 84.342 milliseconds. Short exposure: 50.605 milliseconds. Long exposure: 101.210 milliseconds. With exposure fusion, these images were fused.

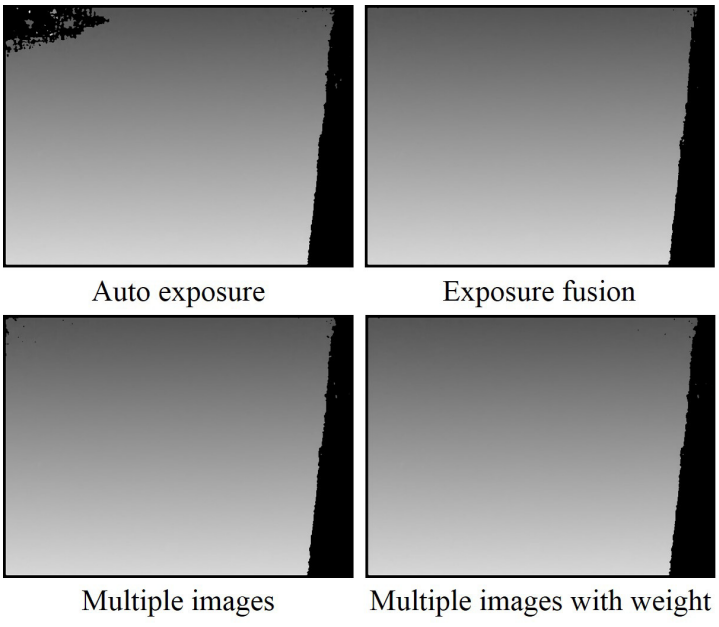


(b) Using the method described in this Chapter, with the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels, the weight image was calculated.

Figure 4.8: The images were acquired in laboratory with the stereo vision camera fixed on a tripod.

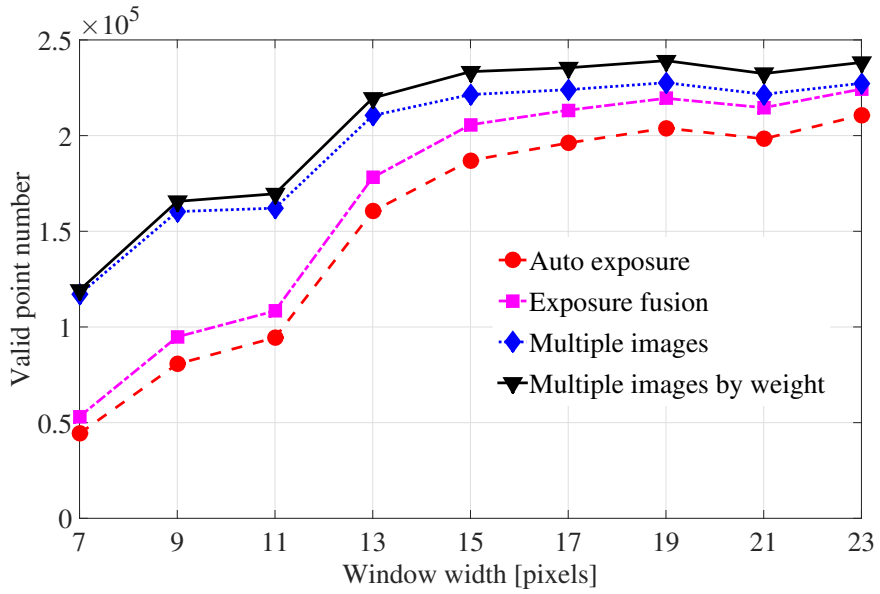


(a) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with local window-based method. The window size is 15x15 pixels.

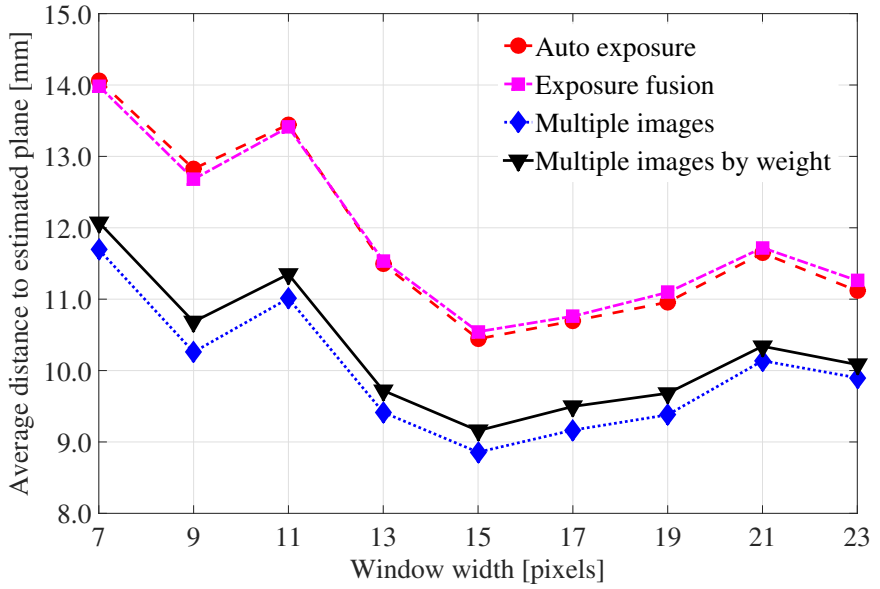


(b) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with SGM. The window size is 11x11 pixels.

Figure 4.9: For the images shown in Figure 4.8, the disparity images were calculated.

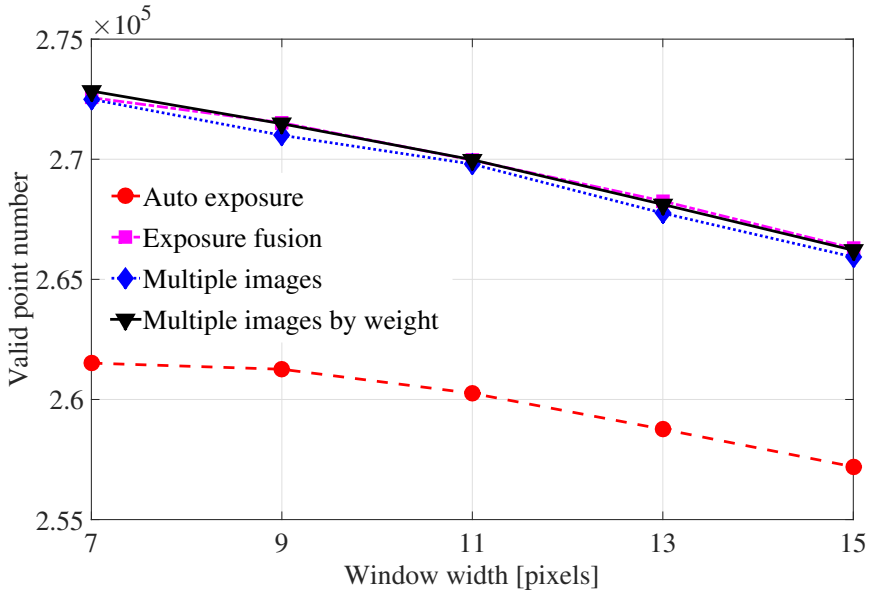


(a) Valid point number N_V.

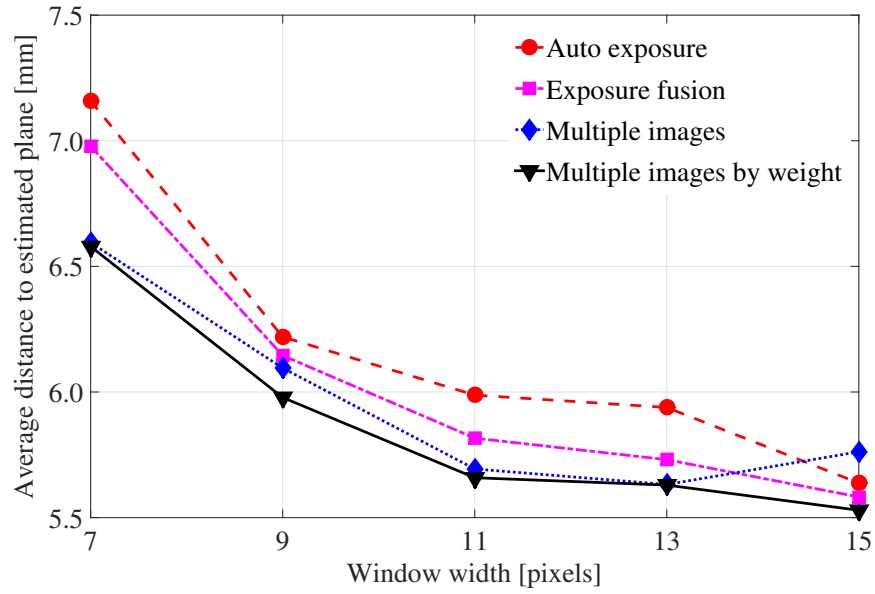


(b) Average distance to the estimated plane d_V.

Figure 4.10: For the images shown in Figure 4.8, using four methods to calculate the matching costs of the matching images, the stereo matching was computed with local window-based method. The window size is changed from 7x7 to 23x23 pixels.



(a) Valid point number N_V.



(b) Average distance to the estimated plane d_V.

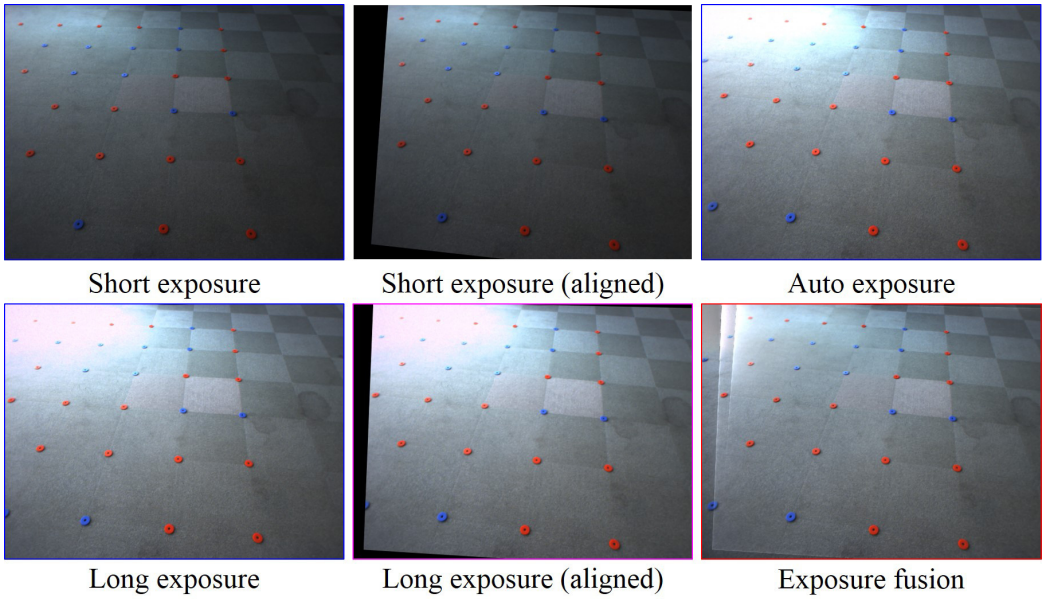
Figure 4.11: For the images shown in Figure 4.8, using four methods to calculate the matching costs of the matching images, the stereo matching was done with SGM. The window size is changed from 7x7 to 15x15 pixels.

part were not calculated. With the window size changed from 7x7 to 23x23 pixels, Figure 4.10 shows the valid point number N_V and the average distance to the estimated plane d_V. It shows that compared with the methods "Auto exposure" and "Exposure fusion", with the methods "Multiple images" and "Multiple images by weight", more valid points can be obtained and the averaged point to estimated plane distance becomes smaller. Especially when the window size is small, the methods "Multiple images" and "Multiple images by weight" performed much better.

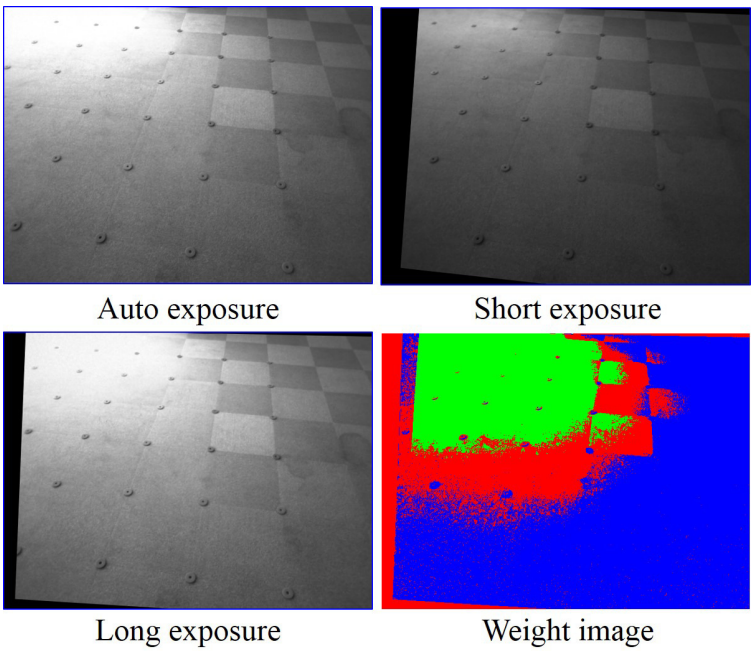
Next, stereo matching was done with SGM. For example, with a window of 11x11 pixels, the disparity images computed with four different methods are shown in Figure 4.9(b). Since it is overexposure in the top left of the image grabbed with auto exposure, the disparity values for the pixels in this part still were not calculated with the method "Auto exposure" even when SGM was used. With the window size changed from 7x7 to 15x15 pixels, Figure 4.11 presents the valid point number N_V and the average distance to the estimated plane d_V. It shows that the N_V of the methods "Exposure fusion", "Multiple images" and "Multiple images by weight" are quite close and bigger than the result of the method "Auto exposure". However, the proposed method "Multiple images by weight" performs best with the smallest d_V.

B. Experiments in laboratory with the stereo vision camera held in the hand

The experiments were done in laboratory with the stereo vision camera held in the hand and the grabbed images of the right camera are shown in Figure 4.12(a). As shown in Figure 4.13(a), the key points of the images grabbed with short and auto exposures were detected with SURF and matched with RANSAC. Similarly, as presented in Figure 4.13(b), the key points of the images grabbed with long and auto exposures were detected with SURF and matched with RANSAC. The images grabbed with short and long exposures were aligned to the image captured with auto exposure and the aligned images are shown in Figure 4.12(a). The image grabbed with

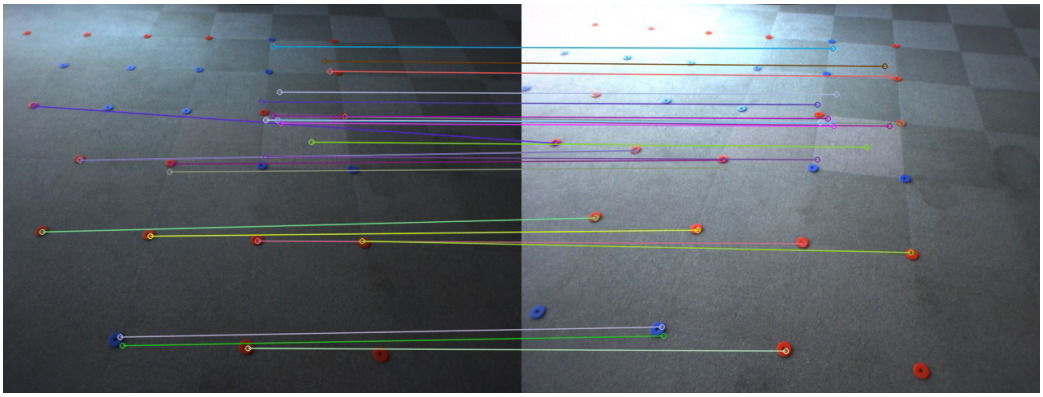


(a) The images were grabbed with different shutter times. Auto exposure: 98.342 milliseconds. Short exposure: 63.405 milliseconds. Long exposure: 126.410 milliseconds. The images grabbed with short and long exposures were aligned to the image captured with auto exposure. With exposure fusion, the image grabbed with auto exposure and the registered images of the photographs captured with short and long exposures were fused.

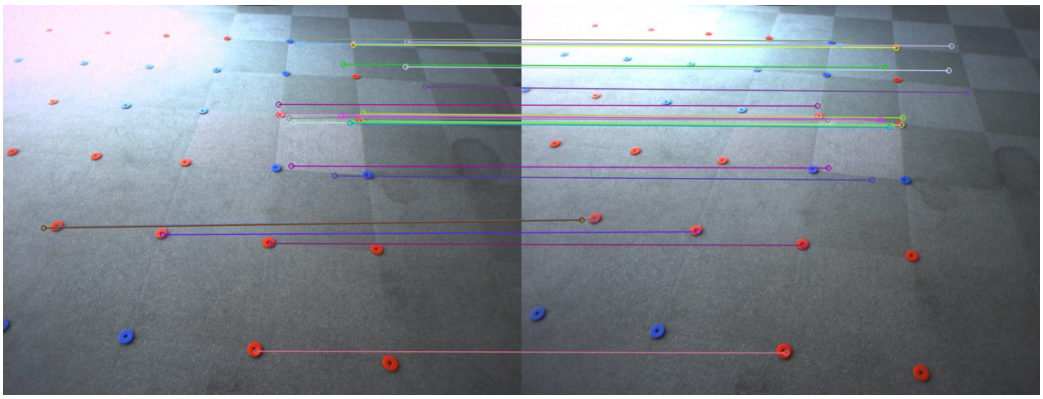


(b) Using the method described in this Chapter, with the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels, the weight image was calculated.

Figure 4.12: The images were acquired in laboratory with the stereo vision camera held in the hand.



(a) The key points were detected and matched between the images with short and auto exposures.

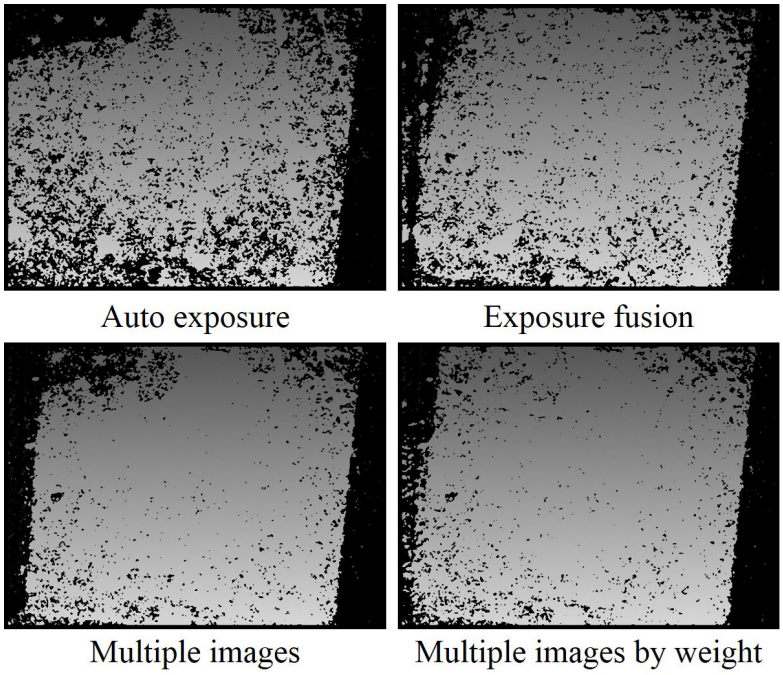


(b) The key points were detected and matched between the images with long and auto exposures.

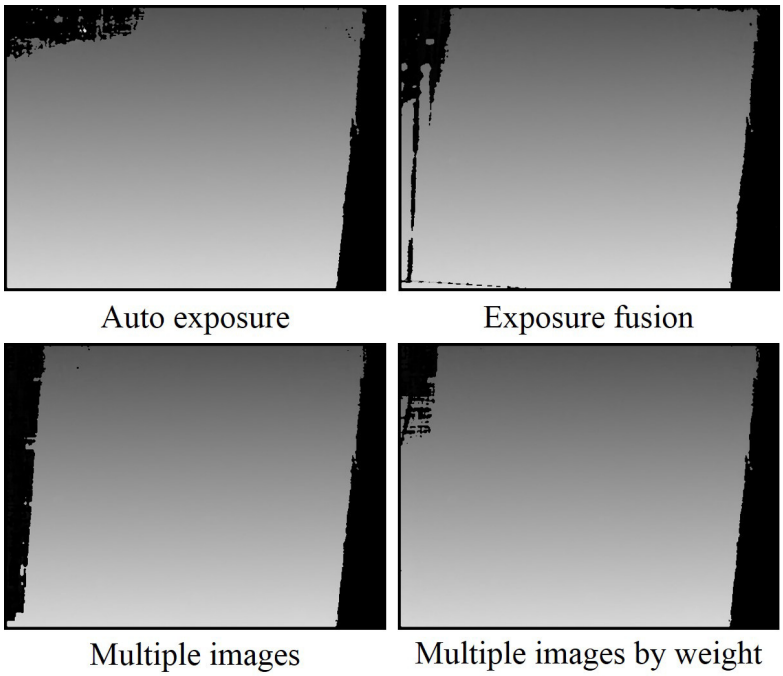
Figure 4.13: With SURF, the key points were detected and matched.

auto exposure and the registered images of the photographs captured with short and long exposures were fused with exposure fusion and the resulting fused image is shown in Figure 4.12(a). The colour images were converted to grayscale images as shown in Figure 4.12(b). As an example, the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels and the resulting weight image is shown in Figure 4.12(b).

First, stereo matching was done with local window-based method. For example, with a window of 15x15 pixels, the disparity images calculated with four different methods are shown in Figure 4.14(a). With the window size changed from 7x7 to 23x23 pixels, Figure 4.15 presents the valid point

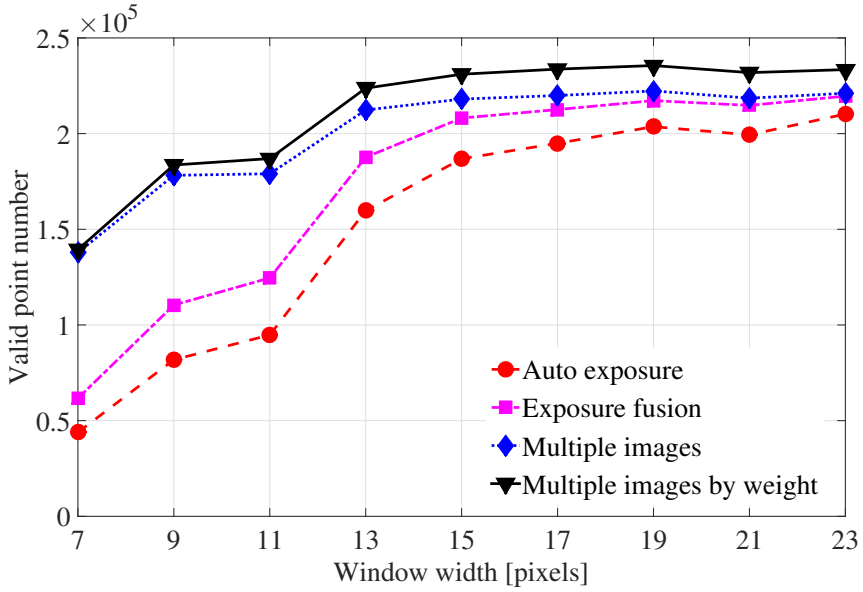


(a) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with local window-based method. The window size is 15x15 pixels.

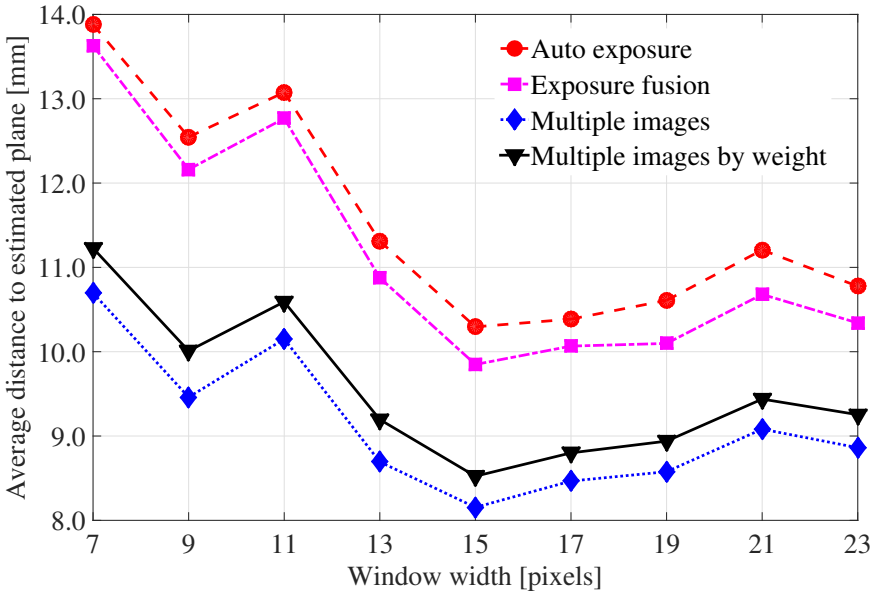


(b) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with SGM. The window size is 11x11 pixels.

Figure 4.14: For the images shown in Figure 4.12, the disparity images were calculated.

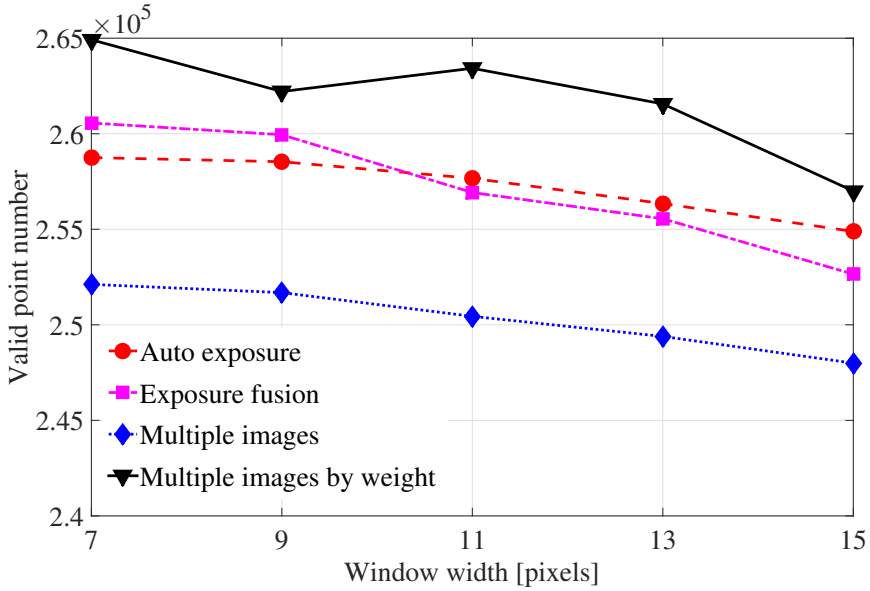


(a) Valid point number N_V.

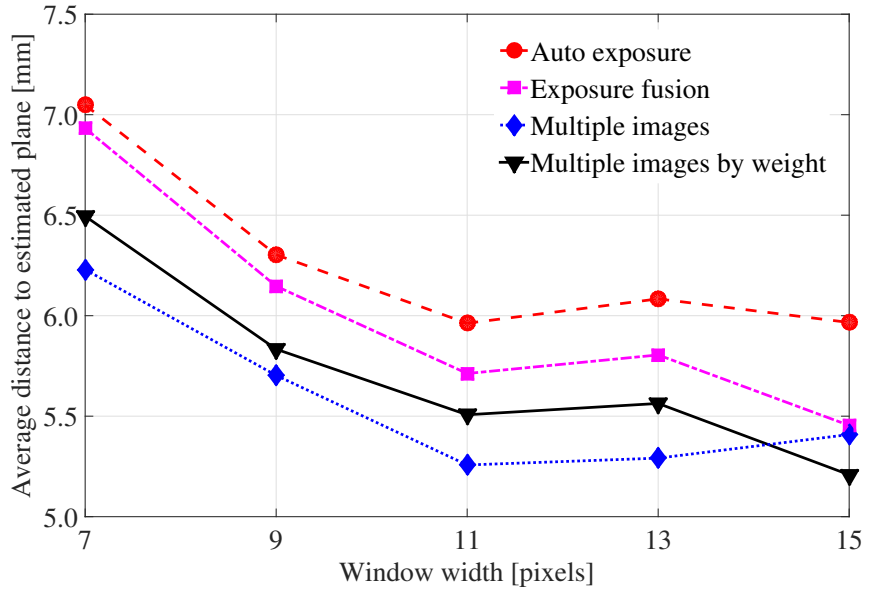


(b) Average distance to the estimated plane d_V.

Figure 4.15: For the images shown in Figure 4.12, using four methods to calculate the matching costs of the matching images, the stereo matching was computed with local window-based method. The window size is changed from 7x7 to 23x23 pixels.



(a) Valid point number N_V.



(b) Average distance to the estimated plane d_V.

Figure 4.16: For the images shown in Figure 4.12, using four methods to calculate the matching costs of the matching images, the stereo matching was done with SGM. The window size is changed from 7x7 to 15x15 pixels.

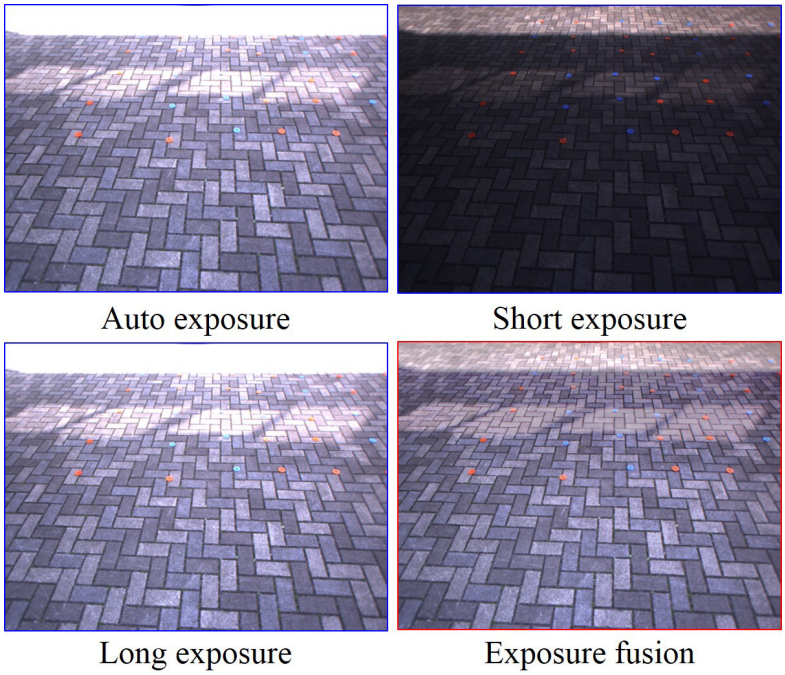
number N_V and the average distance to the estimated plane d_V. It shows that compared with the methods "Auto exposure" and "Exposure fusion", with the methods "Multiple images" and "Multiple images by weight", more valid points are obtained and the averaged point to estimated plane distance becomes smaller. Especially when the window size is small, the methods "Multiple images" and "Multiple images by weight" performed much better.

Next, stereo matching was done with SGM. For example, with a window of 11x11 pixels, the disparity images computed with four different methods are shown in Figure 4.14(b). With the window size changed from 7x7 to 15x15 pixels, Figure 4.16 illustrates the valid point number N_V and the average distance to the estimated plane d_V. It shows that the proposed method "Multiple images by weight" performs best with biggest N_V and its d_V is smaller than the results of the methods "Auto exposure" and "Exposure fusion".

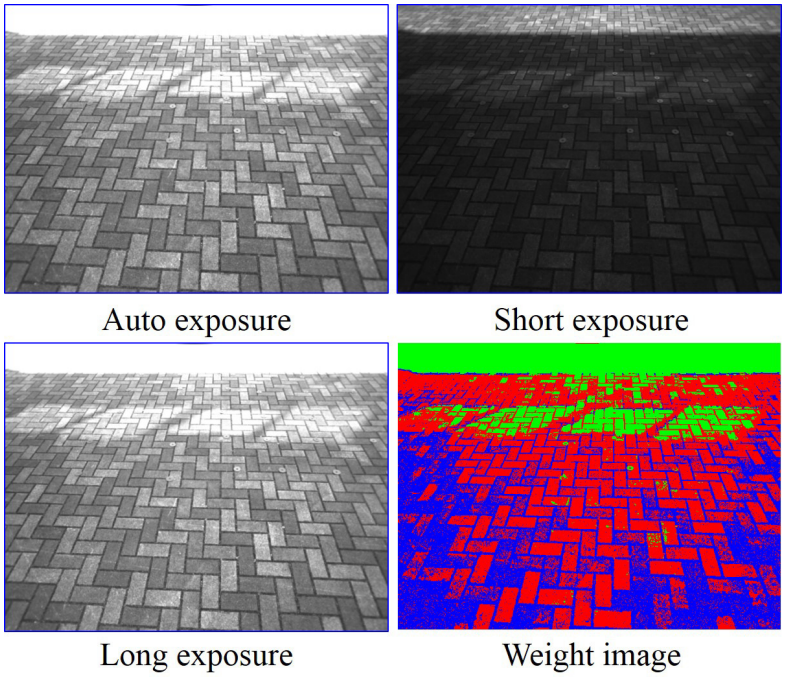
C. Outdoor experiments with the stereo vision camera fixed on a tripod

The experiments were done outdoors with the stereo vision camera fixed on a tripod and the grabbed images of the right camera are shown in Figure 4.17 (a). The image grabbed with auto, short and long exposures were fused with exposure fusion and the resulting fused image is shown in Figure 4.17 (a). The colour images were converted to grayscale images as shown in Figure 4.17(b). As an example, the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels and the resulting weight image is shown in Figure 4.17(b).

First, stereo matching was done with local window-based method. For example, with a window of 15x15 pixels, the disparity images computed with four different methods are shown in Figure 4.18(a). Since it is overexposure in the top of the image grabbed with auto exposure, the disparity values for the pixels in this part were not calculated with the method "Auto exposure". With the window size changed from 7x7 to 23x23 pixels, Figure 4.19 presents the valid point number N_V and the average distance to the estimated plane

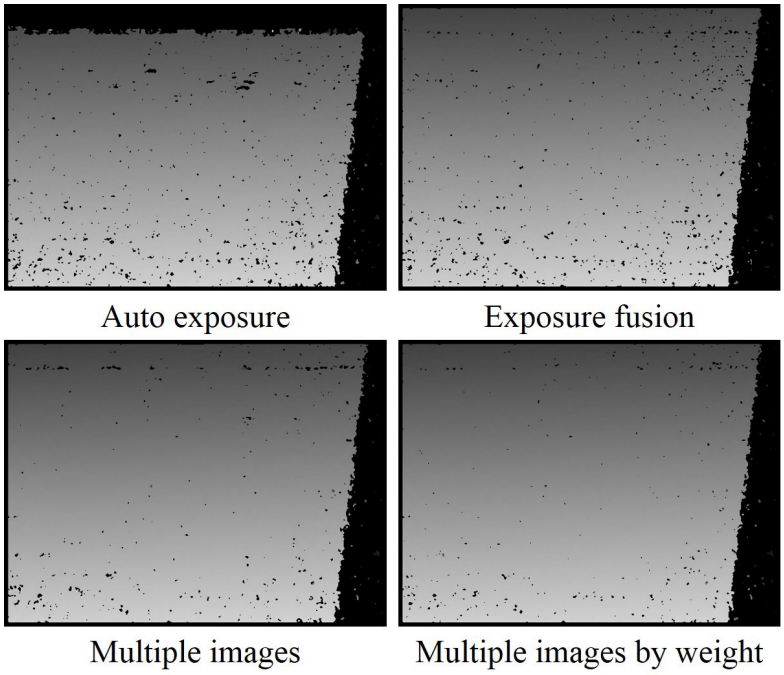


(a) The images were grabbed with different shutter times. Auto exposure: 2.008 milliseconds. Short exposure: 0.482 milliseconds. Long exposure: 2.410 milliseconds. With exposure fusion, these images were fused.

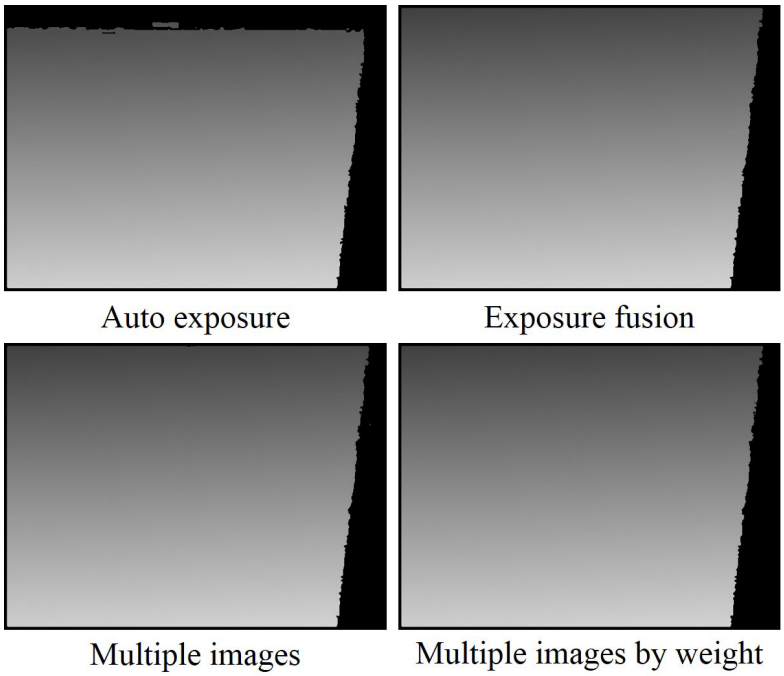


(b) Using the method described in this Chapter, with the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels, the weight image was calculated.

Figure 4.17: The images were acquired outdoors with the stereo vision camera fixed on a tripod.

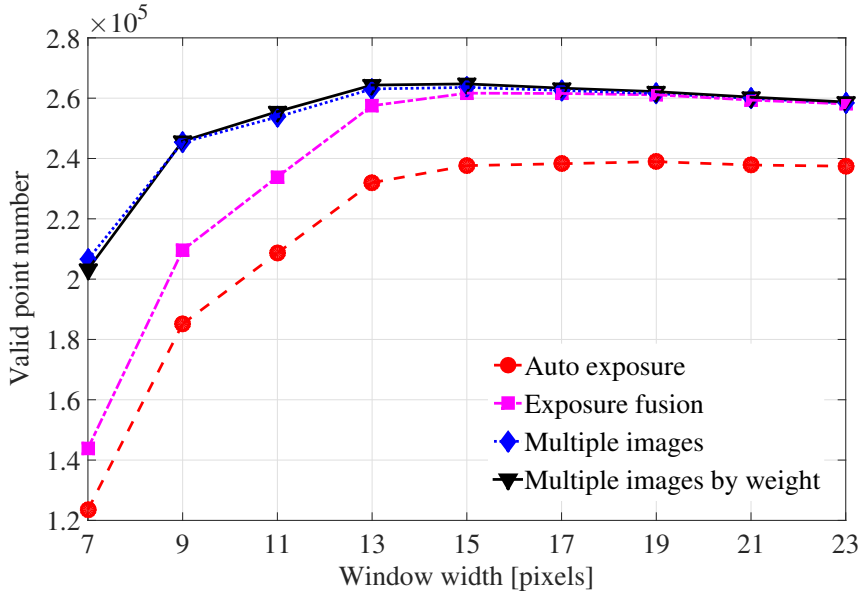


(a) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with local window-based method. The window size is 15x15 pixels.

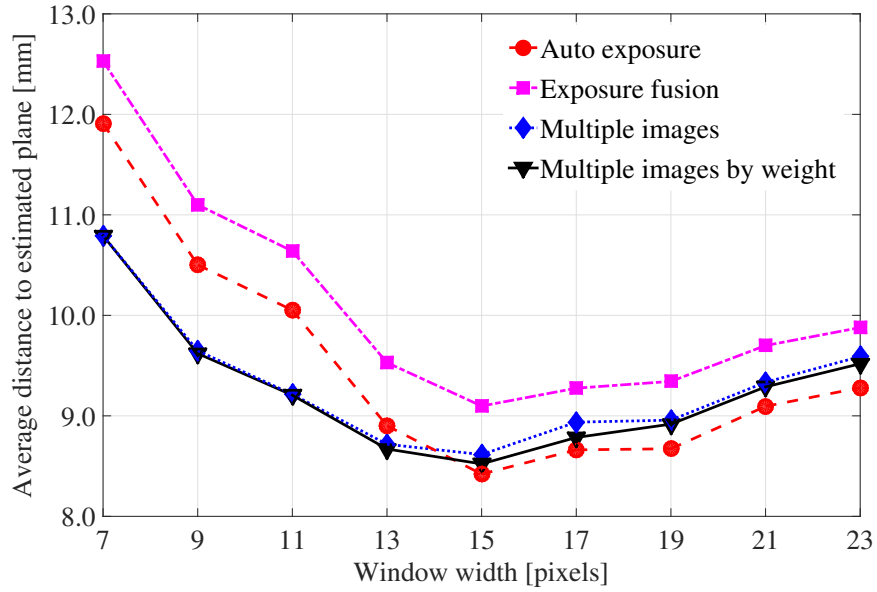


(b) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with SGM. The window size is 11x11 pixels.

Figure 4.18: For the images shown in Figure 4.17, the disparity images were calculated.

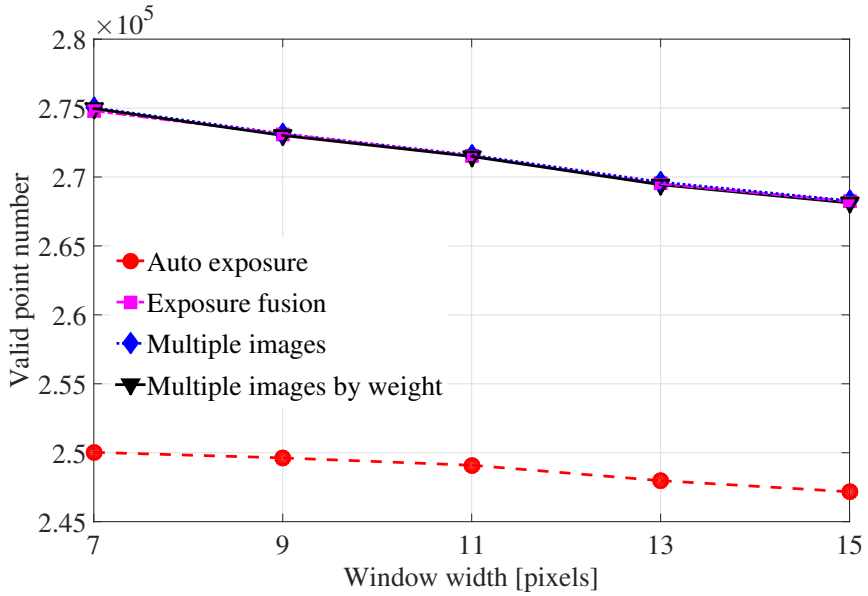


(a) Valid point number N_V.

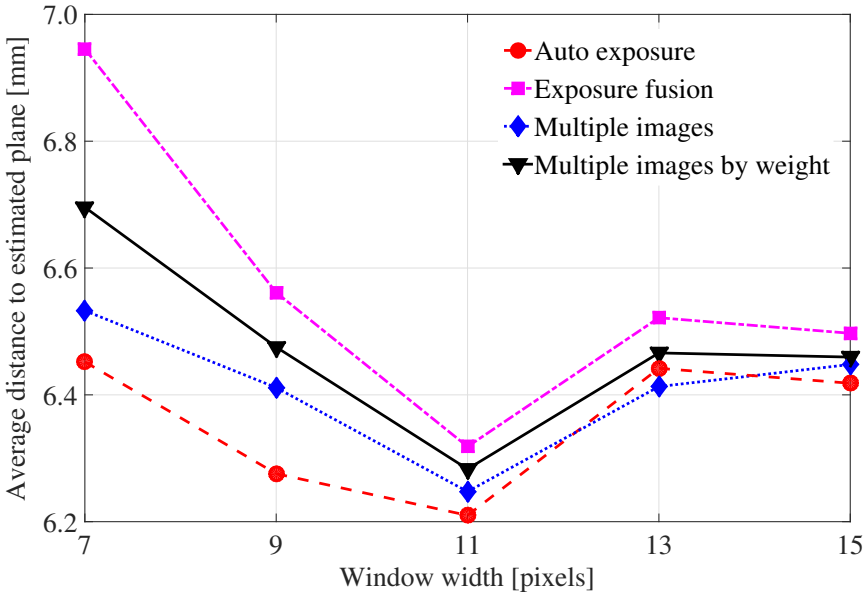


(b) Average distance to the estimated plane d_V.

Figure 4.19: For the images shown in Figure 4.17, using four methods to calculate the matching costs of the matching images, the stereo matching was computed with local window-based method. The window size is changed from 7x7 to 23x23 pixels.



(a) Valid point number N_V.



(b) Average distance to the estimated plane d_V.

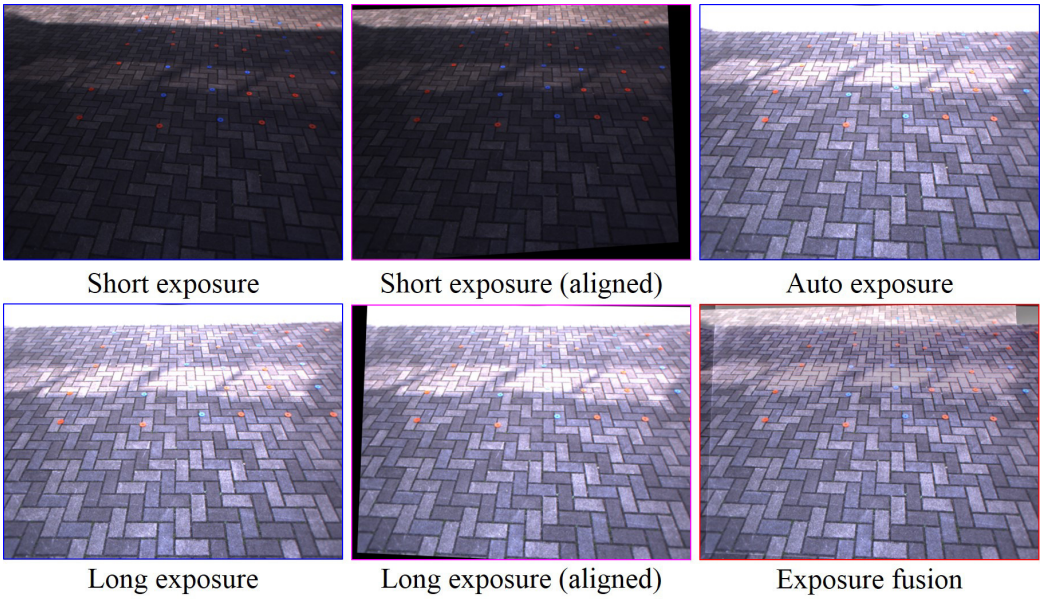
Figure 4.20: For the images shown in Figure 4.17, using four methods to calculate the matching costs of the matching images, the stereo matching was done with SGM. The window size is changed from 7x7 to 15x15 pixels.

d_V. It depicts that compared with the method "Exposure fusion", with the methods "Multiple images" and "Multiple images by weight", more valid points are obtained and the averaged point to estimated plane distance is smaller. Especially when the window size is small, the methods "Multiple images" and "Multiple images by weight" performed much better. It is noticed that in some window sizes it appears that the d_V calculated with the method "Auto exposure" is smallest and the method "Auto exposure" performs "best". However, compared with other methods, its N_V is smallest. Besides, since the disparity values for the pixels in the top of the image were not calculated with the method "Auto exposure" as shown in Figure 4.18(a), the top part of the image was not used to calculated its d_V. For these reasons, the method "Auto exposure" actually performs worst.

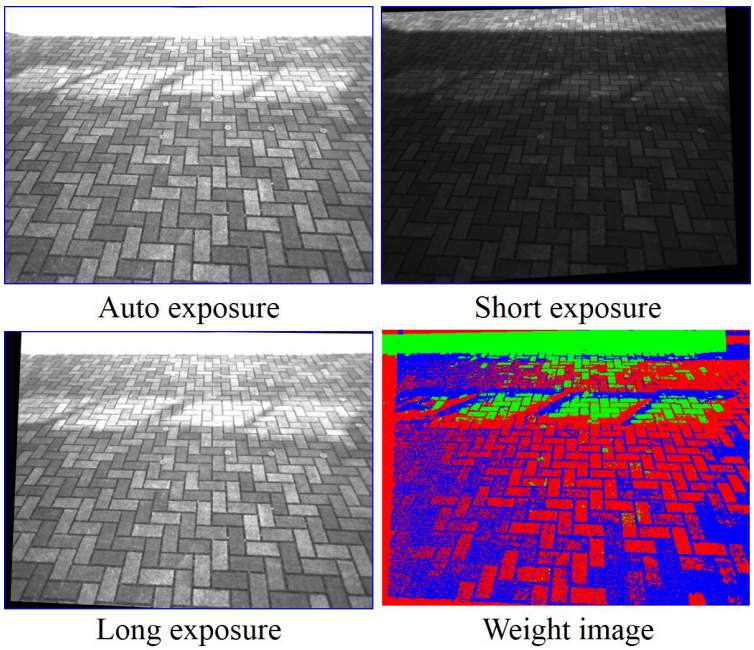
Next, stereo matching was done with SGM. For example, with a window of 11x11 pixels, the disparity images calculated with four different methods are shown in Figure 4.18(b). As it is overexposure in the top of the image grabbed with auto exposure, the disparity values for the pixels in this part still were not calculated with the method "Auto exposure" even when SGM was used. With the window size changed from 7x7 to 15x15 pixels, Figure 4.20 presents the valid point number N_V and the average distance to the estimated plane d_V. It shows that the N_V of the methods "Exposure fusion", "Multiple images" and "Multiple images by weight" are quite close and bigger than the result of the method "Auto exposure". However, the methods "Multiple images" and "Multiple images by weight" perform better than the method "Exposure fusion" with a smaller averaged distances to the estimated plane.

D. Outdoor experiments with the stereo vision camera held in the hand

The experiments were done outdoors with the stereo vision camera held in the hand and the grabbed images of the right camera are shown in Figure 4.21(a). The images grabbed with short and long exposures were aligned to the image captured with auto exposure and the aligned images are shown in Figure 4.21(a). The image grabbed with auto exposure and the regis-

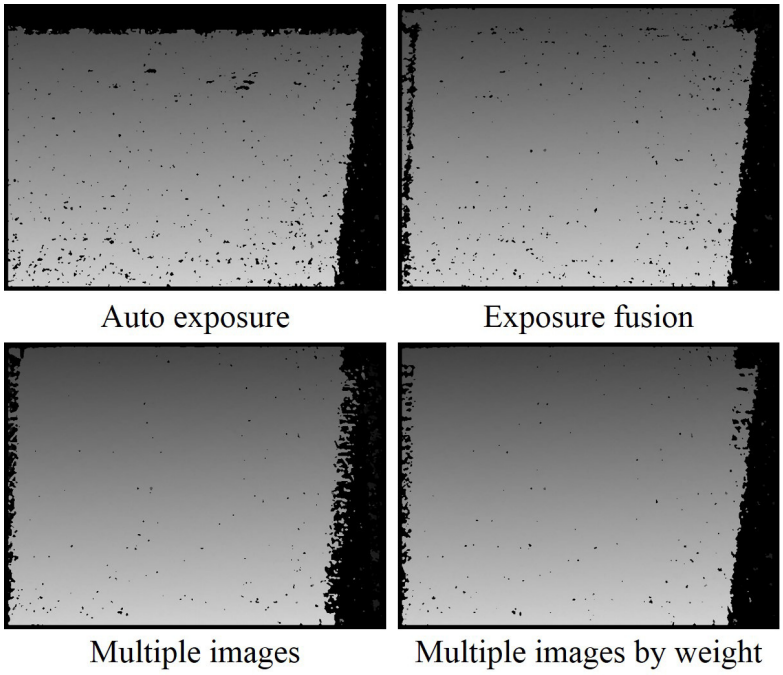


(a) The images were grabbed with different shutter times. Auto exposure: 2.008 milliseconds. Short exposure: 0.482 milliseconds. Long exposure: 2.410 milliseconds. The images grabbed with short and long exposures were aligned to the image captured with auto exposure. With exposure fusion, the image grabbed with auto exposure and the registered images of the photographs captured with short and long exposures were fused.

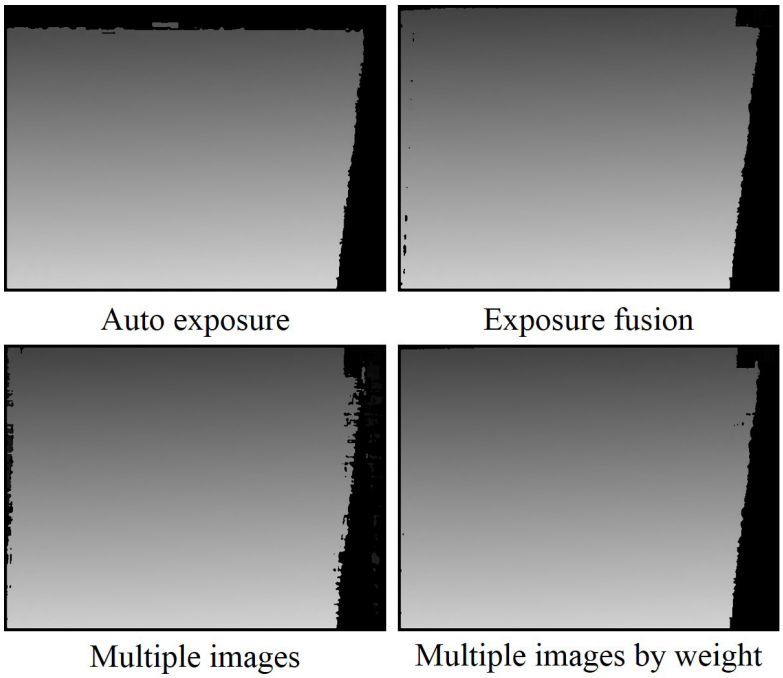


(b) Using the method described in this Chapter, with the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels, the weight image was calculated.

Figure 4.21: The images were acquired outdoors with the stereo vision camera held in the hand.

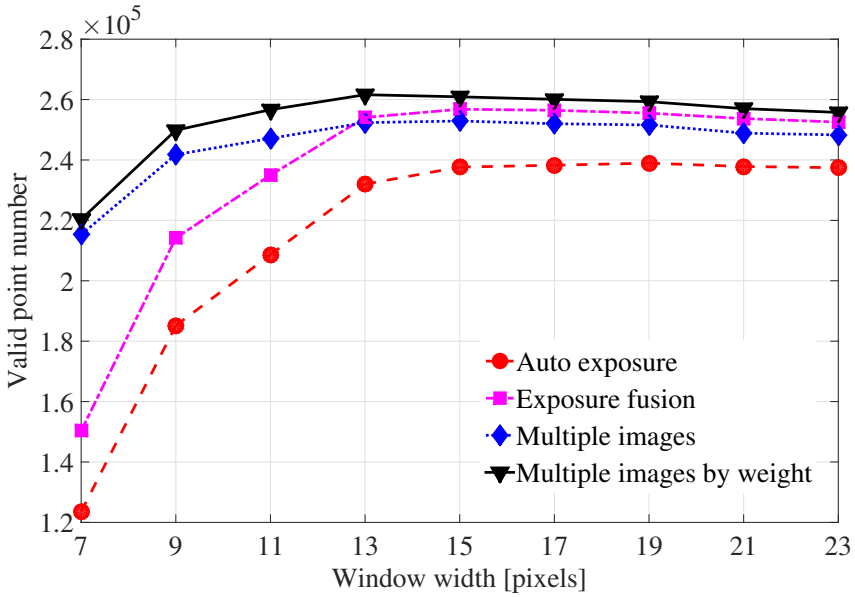


(a) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with local window-based method. The window size is 15x15 pixels.

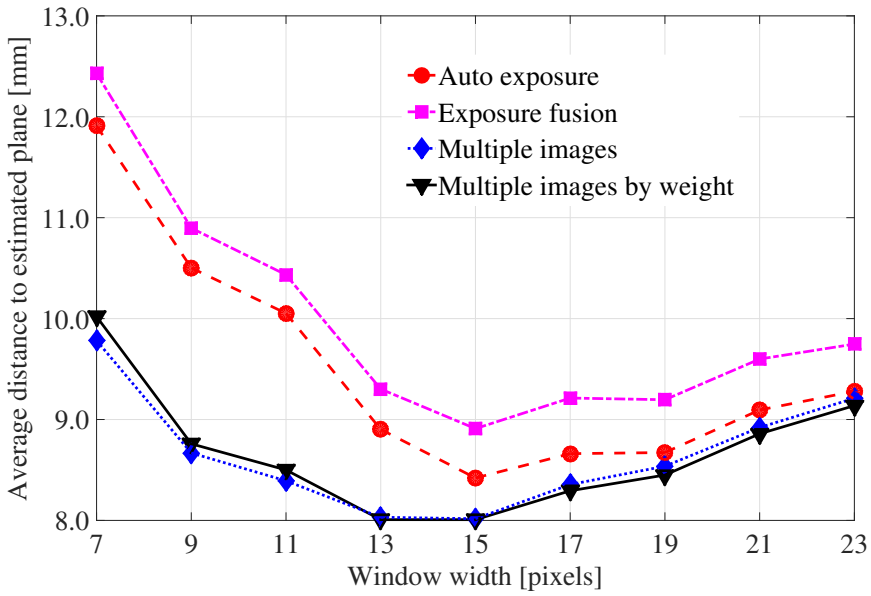


(b) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with SGM. The window size is 11x11 pixels.

Figure 4.22: For the images shown in Figure 4.21, the disparity images were calculated.

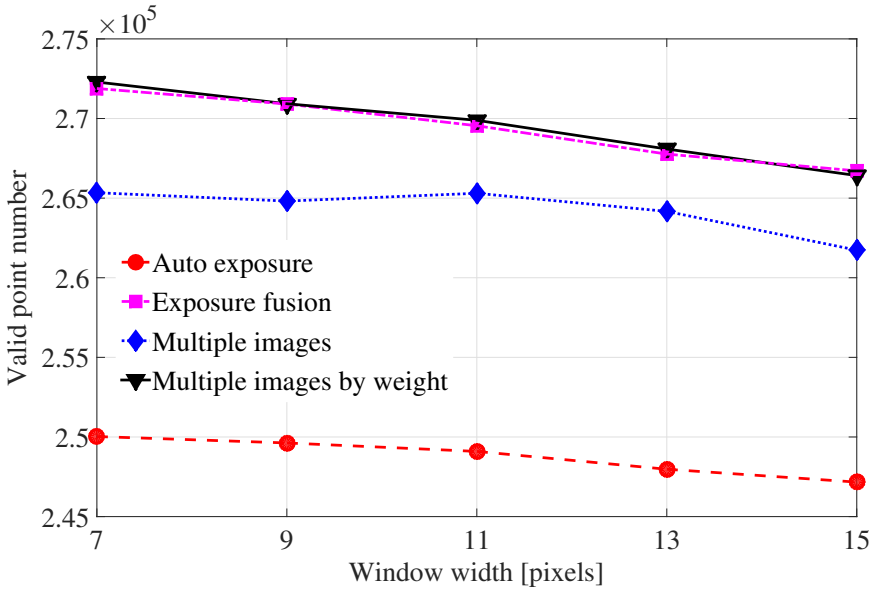


(a) Valid point number N_V.

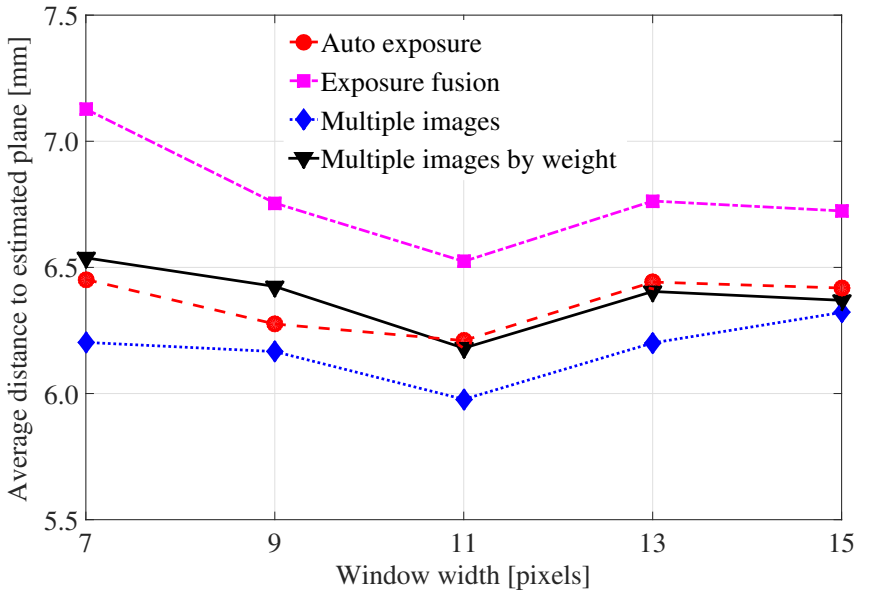


(b) Average distance to the estimated plane d_V.

Figure 4.23: For the images shown in Figure 4.21, using four methods to calculate the matching costs of the matching images, the stereo matching was computed with local window-based method. The window size is changed from 7x7 to 23x23 pixels.



(a) Valid point number N_V.



(b) Average distance to the estimated plane d_V.

Figure 4.24: For the images shown in Figure 4.21, using four methods to calculate the matching costs of the matching images, the stereo matching was done with SGM. The window size is changed from 7x7 to 15x15 pixels.

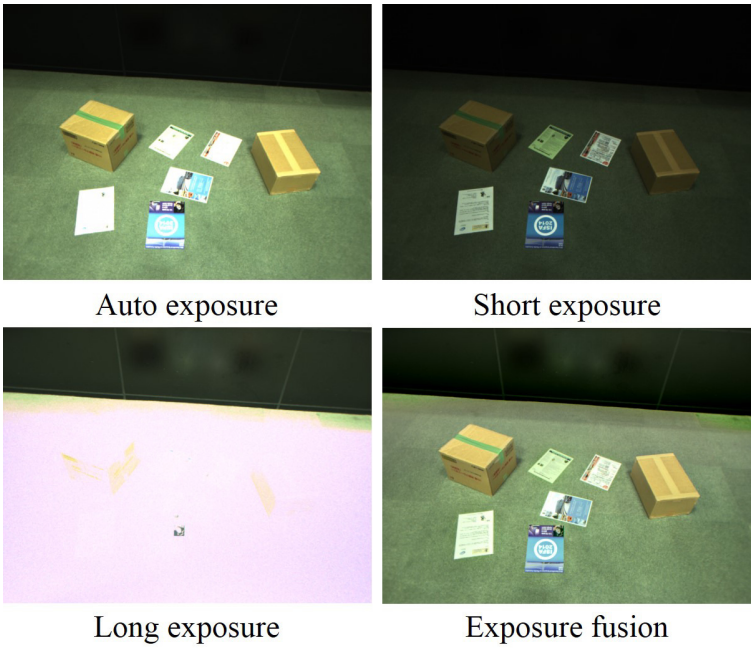
tered images of the photographs captured with short and long exposures were fused with exposure fusion and the resulting fused image is shown in Figure 4.21(a). The colour images were converted to grayscale images as shown in Figure 4.21(b). As an example, the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels and the weight image is shown in Figure 4.21(b).

First, stereo matching was done with local window-based method. For example, with a window of 15x15 pixels, the disparity images calculated with four different methods are shown in Figure 4.22(a). With the window size changed from 7x7 to 23x23 pixels, Figure 4.23 shows the valid point number N_V and the average distance to the estimated plane d_V. It illustrates that with the proposed method "Multiple images by weight", more valid points can be obtained and the averaged point to estimated plane distance becomes smaller compared with the methods "Auto exposure" and "Exposure fusion". Especially when the window size is small, the proposed method "Multiple images by weight" performed much better.

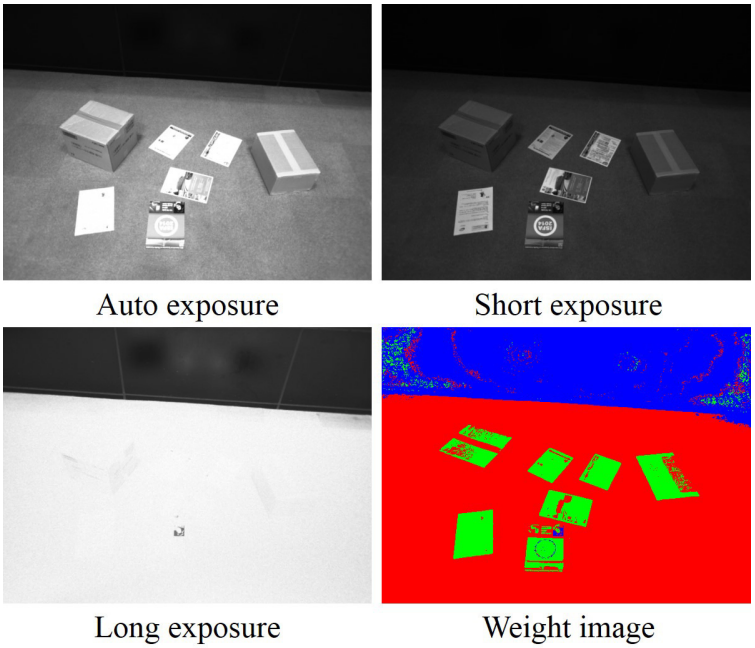
Next, stereo matching was done with SGM. For example, with a window of 11x11 pixels, the disparity images computed with four different methods are shown in Figure 4.22(b). With the window size changed from 7x7 to 15x15 pixels, Figure 4.24 presents the valid point number N_V and the average distance to the estimated plane d_V. It shows that the N_V of the methods "Exposure fusion" and "Multiple images by weight" are quite close and bigger than the results of the methods "Auto exposure" and "Multiple images". However, the proposed method "Multiple images by weight" performs better than the method "Exposure fusion" with a smaller d_V.

4.2.3 Experiment of environment perception

The experiments were done for the environment perception and the grabbed images of the right camera are shown in Figure 4.25(a). The image grabbed with auto, short and long exposures were fused with exposure fusion and the resulting fused image is shown in Figure 4.25 (a). The colour images were converted to grayscale images as shown in Figure 4.25(b). As an example,

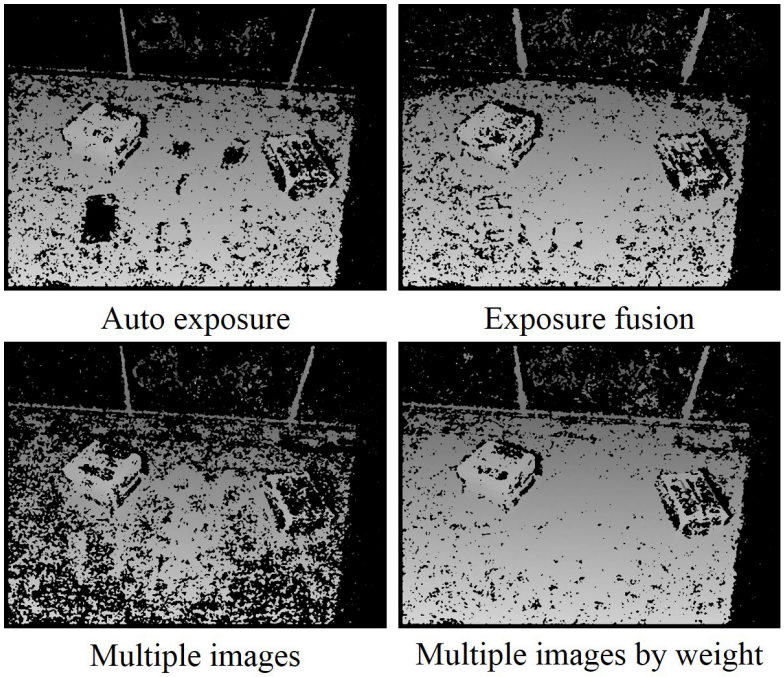


(a) The images were grabbed with different shutter times. Auto mode: 164.842 milliseconds. Short exposure: 59.343 milliseconds. Long exposure: 708.819 milliseconds. With exposure fusion, these images were fused.

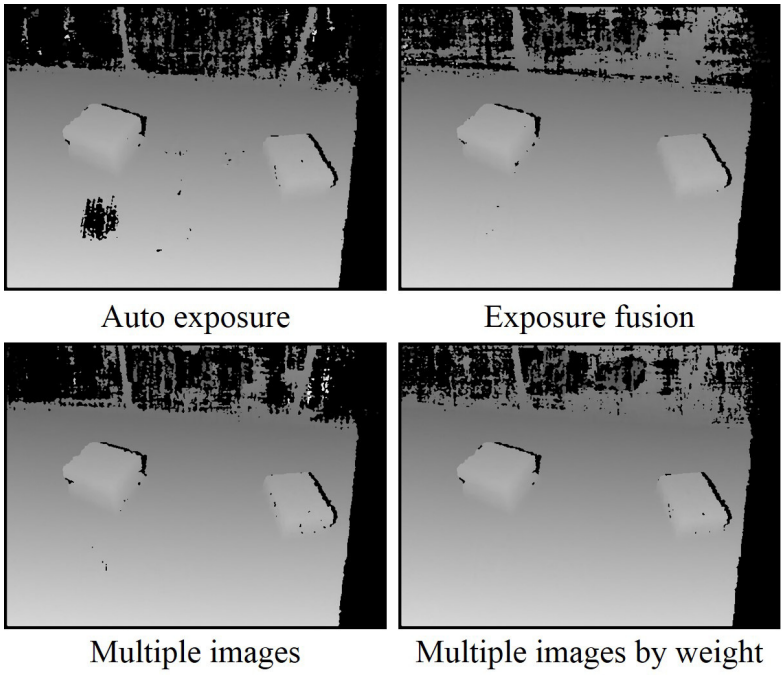


(b) Using the method described in this Chapter, with the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels, the weight image was calculated.

Figure 4.25: The images were acquired for environment perception.



(a) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with local window-based method. The window size is 15x15 pixels.



(b) Using four methods to calculate the matching costs of the matching images, the disparity images were calculated with SGM. The window size is 11x11 pixels.

Figure 4.26: For the images shown in Figure 4.25, the disparity images were calculated.

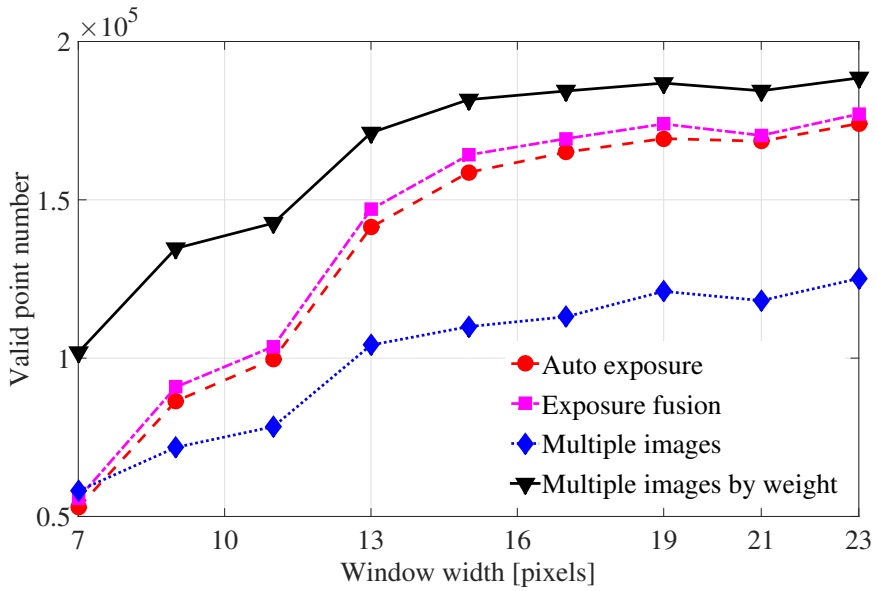


Figure 4.27: For the images shown in Figure 4.25, using four methods to calculate the matching costs of the matching images, the valid point numbers were presented and the stereo matching was computed with local window-based method. The window size is changed from 7x7 to 23x23 pixels.

the intensity diversity weight of each pixel was calculated with a window of 15x15 pixels and the resulting weight image is shown in Figure 4.25(b).

First, stereo matching was done with local window-based method. For example, with a window of 15x15 pixels, the disparity images calculated with four different methods are shown in Figure 4.26(a). With the window size changed from 7x7 to 23x23 pixels, Figure 4.27 shows the valid point number. It illustrates that with proposed method "Multiple images by weight", more valid points can be obtained, especially when the window size is small.

Next, stereo matching was done with SGM. For example, with a window of 11x11 pixels, the disparity images computed with four different methods are shown in Figure 4.26(b). With the window size changed from 7x7 to 15x15 pixels, the valid point number is shown in Figure 4.28. From Figure 4.28, it shows that the proposed method "Multiple images by weight" performs better than other methods with more valid points.

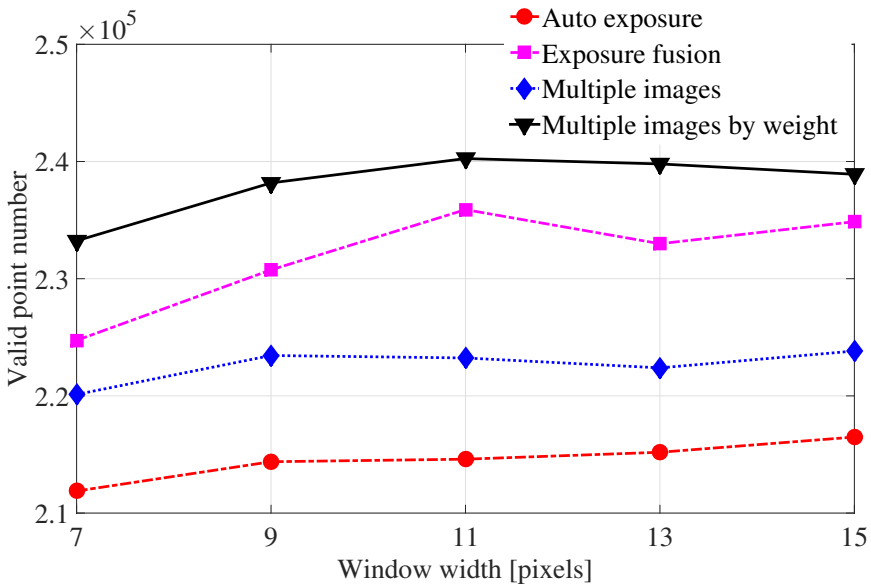


Figure 4.28: For the images shown in Figure 4.25, using four methods to calculate the matching costs of the matching images, the valid point numbers were presented and the stereo matching was done with SGM. The window size is changed from 7x7 to 15x15 pixels.

4.2.4 Discussion

Compared to the method "Exposure fusion", for each pixel of the matching image, the local information in its local window acquired from the images grabbed with auto, short and long exposures is better retained when the matching cost is aggregated with the methods "Multiple images" and "Multiple images by weight". As an example, for the pixel $\boldsymbol{p}(50, 125)$ shown in Figure 4.29, with a window of 15x15 pixels, its matching costs computed with local window-based method and SGM are shown in Figure 4.30(a) and Figure 4.30(b) respectively. It is noticed that compared with the methods "Auto exposure" and "Exposure fusion", the matching costs computed with the methods "Multiple images" and "Multiple images by weight" show distinct minima. With the proposed method "Multiple images by weight", the matching cost value obtained from the pixel which is well exposed and has significantly different intensity values in its local window becomes dominant

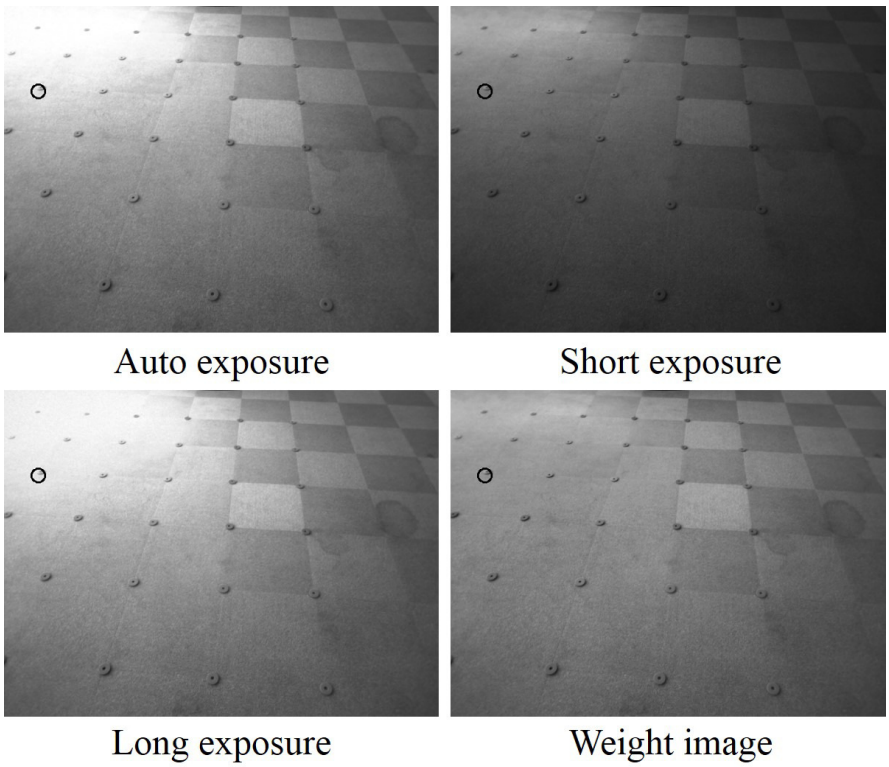
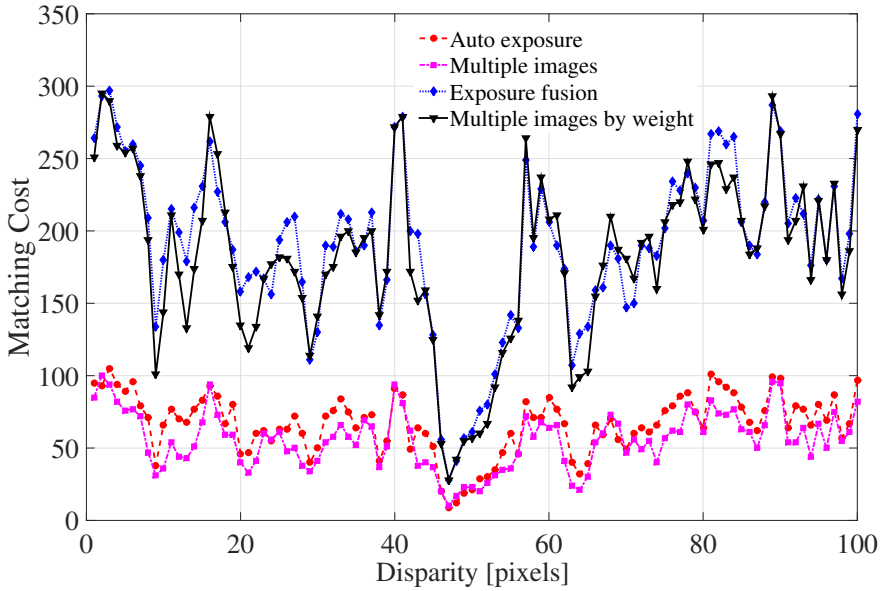


Figure 4.29: For the images shown in Figure 4.8, the pixel $\boldsymbol{p}(50, 125)$ was circled.

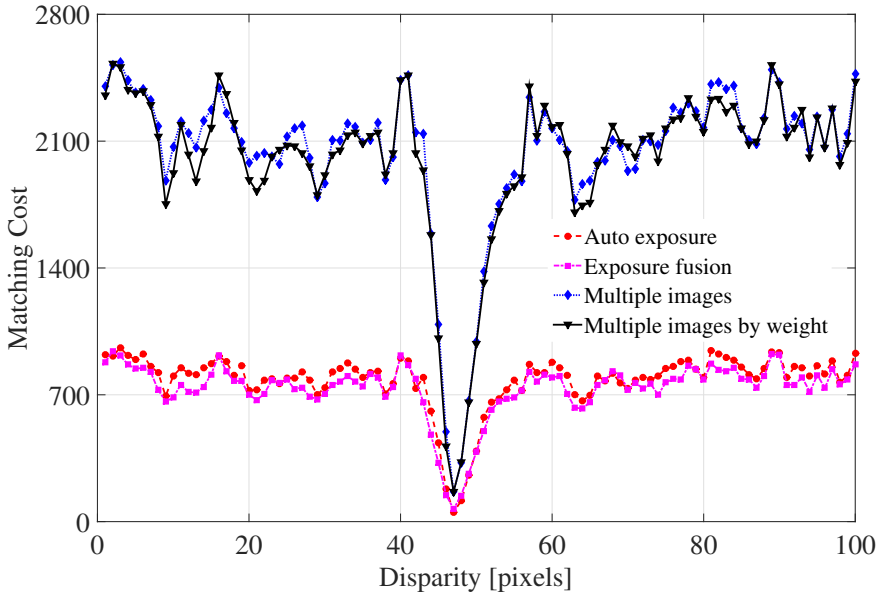
and the useful information in the local window is well retained. For this reason, the experimental results show that compared with the methods "Auto exposure" and "Exposure fusion", the proposed method consistently allowed more valid points to be obtained and the 3D terrain model could be built more accurately. Since the matching cost is not smoothed when it is computed with the method "Multiple images", the method "Multiple images" performed worse than the proposed method with less valid points.

4.3 Conclusion

In order to apply stereo vision techniques in field robotics to acquire 3D terrain maps in extreme light conditions, a series of photographs are taken



(a) Using four methods, the matching costs were computed with window-based method.



(b) Using four methods, the matching costs were computed with SGM.

Figure 4.30: For the pixel $\boldsymbol{p}(50, 125)$ shown in Figure 4.29, its matching costs were calculated with a window of 15x15 pixels.

with multiple exposures. Since it is possible that the camera is moved when the images are grabbed with multiple exposures, the images acquired with short and long exposures are aligned to the image grabbed with auto exposure. A stereo matching algorithm, the matching costs of the images grabbed with multiple exposures are directly summed by weight, is proposed in this Chapter. Compared to the traditional methods such as "Exposure fusion", with the proposed method, it is not needed to fuse the images grabbed with multiple exposures, and for each pixel of the matching image, the local information in its local window acquired from the images grabbed with multiple exposures could be better retained. Experiments were done in laboratory and outdoors with a stereo vision camera fixed on a tripod and held in the hand, and the stereo matching were done with a local window-based method and SGM. Through the experiments, it was verified that compared with the methods "Auto exposure", "Exposure fusion" and "Multiple images", the proposed method consistently allowed more valid points to be obtained and the 3D terrain model could be built more accurately. Especially when the local window-based method was used, compared to other methods, the proposed method performed much better compared to the traditional methods.

Chapter 5

Conclusion and future work

5.1 Conclusion

Stereo vision camera is widely used in the robotics area for terrain mapping, object detection, object classification, navigation and so on. This thesis presents the problems that the Humanitarian Demining Robot Gryphon meets in the field when a stereo vision camera is used for terrain mapping: accurate Kinematic calibration and Hand-Eye calibration and ensuring the 3D terrain model still could be well acquired with a stereo vision camera even in the extreme light conditions. These problems are also the basic and important issues when applying the stereo vision camera in the field and this thesis focuses on them.

First, the Pure Rotation with Joint Angle Constraint Method, which relies on a sequence of pure rotations of the manipulator links, while using a stereo vision camera to track an internal fixed reference point on the manipulator's tip and an external arbitrary reference point fixed in the robot coordinate frame, is proposed and introduced in detail. With the proposed method, the kinematic calibration and hand-eye calibration can be performed simultaneously. The Humanitarian Demining Robot system was used as a platform to verify it. The proposed calibration method was compared with other calibration methods and verified by experiments using an external high-precision hardware. Through experiments, it is verified that the proposed

calibration method performs better than the traditional pure rotation calibration method (the joint angle constraint is not used to fit the circular arc) and the commonly used calibration method (nonlinear least squares estimation with LMA) for the reason that with the proposed calibration method a short circular arc can be fitted more accurately because the joint angle is added as a constraint to fit the circular arc. With the proposed calibration method, the joint initial angles of the manipulator are calibrated with the reference point position error at the manipulator's tip smaller than 11.0 mm and the stereo vision camera can map the terrain with the position error smaller than 16.0 mm which meet the error requirement of the Humanitarian Demining Robot system. The calibration method described in this thesis can be used in other robot platforms to do kinematic calibration and hand-eye calibration.

Next, the Pure Rotation with Joint Angle Constraint Method is improved and the circular arc is directly fitted in 3D space with joint angle constraint. Through the simulation results and experimental results, it is proved that compared to the traditional methods, in 3D space a circular arc can be fitted more precisely with the proposed method, especially when a short arc is fitted and/or the measured data is noisy. The pure rotation calibration method was used to do the Hand-Eye calibration for a Pan-Tilt-Camera system, and it was verified that the proposed methods performed better than the traditional methods. The proposed method of directly fitting a circular arc in 3D space can be used to other application when the actual joint angle can be measured.

Finally, a stereo matching algorithm, which is directly done using the images grabbed with multiple exposures, is proposed and presented in detail. In order to apply the stereo vision techniques in field to acquire 3D terrain maps in extreme light conditions, a series of photographs are taken through multiple exposures. Since it is possible that the camera is moved when the images are grabbed with multiple exposures, the images acquired with short and long exposures are aligned to the image grabbed with auto exposure. The matching costs of the resulting registered images and the image grabbed with auto exposure are directly summed by weight. Compared with the traditional methods such as "Exposure fusion", with the proposed

method, it is not needed to fuse the images grabbed with multiple exposures. When the matching cost is computed with the proposed method, for each pixel of the matching image, the local information in its local window acquired from the images grabbed with multiple exposures could be better retained. Experiments were done in laboratory and outdoors with a stereo vision camera fixed on a tripod and held in the hand, and stereo matching were done with a local window-based method and SGM. Through the experiments, it is verified that compared to the traditional methods such as "Auto exposure" and "Exposure fusion", the proposed method consistently allowed more valid points to be obtained and the 3D terrain model could be built more accurately. Especially when the local window-based method was used, the proposed method performed much better than the traditional methods.

Field experiments are planned to be conducted with the Gryphon system in Angola in the near future to further evaluate the proposed methods. The proposed methods can be used in other robot platforms.

5.2 Future work

In the future, there are some issues that could be studied to further improve the performance of the Humanitarian Demining Robot system. First, *the On-line Calibration could be used for the Kinematic calibration and Hand-Eye calibration.* As presented in Figure 5.1, when a stereo vision camera is moved from one position to the next position, the 3D maps are acquired in these two positions. Through the point clouds in these two different positions, the transfer matrix between them, $\boldsymbol{R}$ and $\boldsymbol{T}$, could be obtained and the point clouds are fused. As shown in Figure 5.2, with the transfer matrix $\boldsymbol{R}$ and $\boldsymbol{T}$, the parameters of the Kinematic calibration and Hand-Eye calibration could be calibrated and checked on-line.

Next, *the Humanitarian Demining Robot system could work as an autonomous system.* Currently, it is remotely controlled. With the stereo vision camera, Inertial Measurement Unit (IMU), Global Positioning System (GPS) and so on, for the Humanitarian Demining Robot system, the local and global maps could be built. With the global 3D map, the robot could

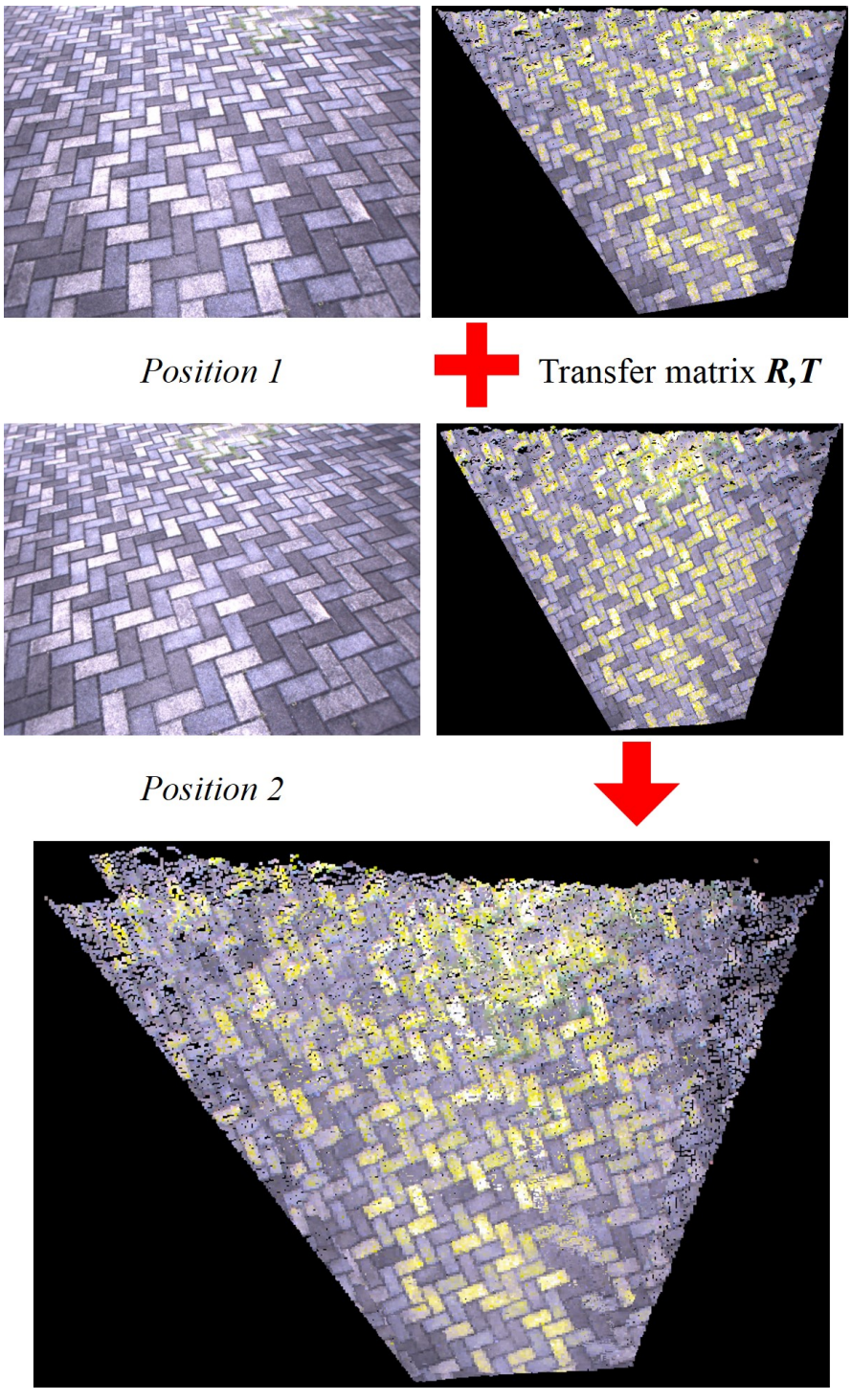


Figure 5.1: 3D maps are acquired in two different positions. Through the point clouds in these two different positions, the transfer matrix between them, $\boldsymbol{R}$ and $\boldsymbol{T}$, could be obtained.

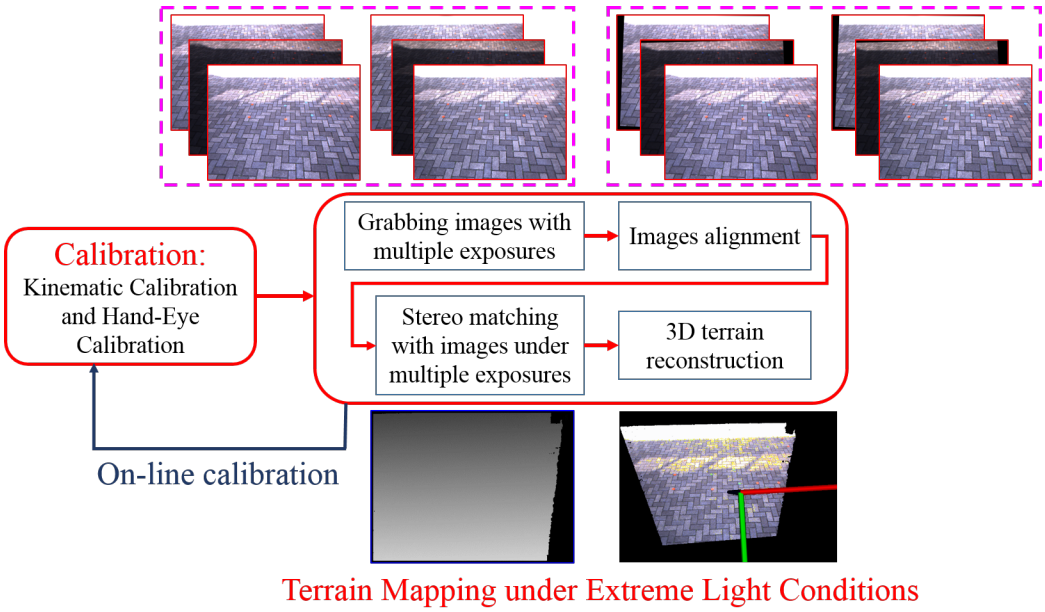


Figure 5.2: On-line Calibration.

automatically scan the ground. Besides, with the global 3D map obtained by one robot, many robots could be placed at different positions and they could automatically scan the ground at the same time.

Finally, *the sensor fusion of stereo vision camera and laser range finder could be applied for terrain mapping.* Currently, the stereo vision camera is used for terrain mapping. However, the laser range finder is still considered as an options. With the sensor fusion of stereo vision camera and laser range finder, a 3D map could be generated more robustly.

References

[1] Edwardo F. Fukushima, Marc Freese, Toshiaki Matsuzawa, Takatoshi Aibara and Shigeo Hirose. "Humanitarian Demining Robot Gryphon Current Status and an Objective Evaluation". *International Journal on Smart Sensing and Intelligent Systems*, vol.1, no.3, pp.735-753, 2008.

[2] Stefan Hrabar. "An evaluation of stereo and laser-based range sensing for rotorcraft unmanned aerial vehicle obstacle avoidance". *Journal of Field Robotics*, vol.29, Issue 2, pp.215-239, 2012.

[3] Bruno Steux and Oussama EL Hamzaoui. "tinySLAM: A SLAM algorithm in less than 200 lines C-language program". *Control Automation Robotics and Vision (ICARCV) 2010 11th International Conference on*, pp.1975-1979, Singapore, December, 2010.

[4] Markus Achtelik, Abraham Bachrach, Ruijie He, Samuel Prentice and Nicholas Roy. "Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments". *In Proceedings of the SPIE Unmanned Systems Technology XI*, vol.7332, Orlando, FL, USA, 2009.

[5] Zvi S. Roth, Benjamin W. Mooring and Bahram Ravani. "An overview of robot calibration". *IEEE Journal of Robotics and Automation*, vol.RA-3, no.5, pp.377-385, 1987.

[6] John M. Hollerbach and Charles W. Wampler "The Calibration Index and Taxonomy for Robot Kinematic Calibration Methods". *The International Journal of Robotics Research*, vol.15, Issue 6, pp. 573-591, 1996.

[7] Benjamin W. Mooring, Zvi S. Roth, Morris R. Driels. "Fundamentals of Manipulator Calibration". *John Wiley & Sons, Inc.*, New York, 1991.

[8] Milan Ikits and John M. Hollerbach. "Kinematic Calibration Using a Plane Constraint". *IEEE International Conference on Robotics and Automation*, pp.3191-3196, Albuquerque, New Mexico, 1997.

[9] Joon Hyun Jang, Soo Hyun Kim and Yoon Keun Kwak. "Calibration of geometric and non-geometric errors of an industrial robot". *Robotica*, vol.19, pp.311-321, 2001.

[10] Seiji Aoyagi, Atsushi Kohama, Yasutaka Nakata, Yuki Hayano, Masato Suzuki. "Improvement of Robot Accuracy by Calibrating Kinematic Model Using a Laser Tracking System -Compensation of Non-Geometric Errors Using Neural Networks and Selection of Optimal Measuring Points Using Genetic Algorithm-". *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp.5660-5665, October, 2010.

[11] Tobias Kastner, Thomas Rofer and Tim Laue. "Automatic Robot Calibration for the NAO". *RoboCup 2014: Robot World Cup XVIII Lecture Notes in Computer Science*, vol.8992, pp.233-244, 2015.

[12] Henry W. Stone and Arthur C. Sanderson. "A prototype arm signature identification system". *Robotics and Automation. Proceedings. 1987 IEEE International Conference on*, vol.4, pp.175-182, 1987.

[13] Wyatt S. Newman, Craig E. Birkhimer and Robert J. Horning. "Calibration of a Motoman P8 robot based on laser tracking". *Robotics and Automation, 2000. Proceedings of the IEEE International Conference on Robotics and Automation*, pp.3597-3602, 2000.

[14] Richard I. Hartley. "Self-calibration from multiple views with a rotating camera". *Third European Conference on Computer Vision, 1994 Proceedings*, vol.800, pp.471-478, May, 1994.

[15] G. P. Stein. "Accurate internal camera calibration using rotation, with analysis of sources of error". *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pp.230-236, June, 1995.

[16] Jorge Santolaria, Javier Conte, Marcos Pueo and Carlos Javierre. "Rotation error modeling and identification for robot kinematic calibration by circle point method". *Metrol. Meas. Syst.*, vol.XXI, no.1, pp.85-98, 2014.

[17] Radu Horaud and Fadi Dornaika. "Hand-Eye Calibration". *International Journal of Robotics Research*, vol.14, no.3, pp.195-210, June, 1995.

[18] Klaus H. Strobl and Gerd Hirzinger. "Optimal Hand-Eye Calibration". *Proceedings of the 2006 IEEE/RSJ, International Conference on Intelligent Robots and Systems*, pp.4647-4653, October, 2006.

[19] Yiu Cheung Shiu and Shaheen Ahmad. "Calibration of Wrist Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form AX = XB". *Robotics and Automation, IEEE Transactions on*, vol.5, no.1, pp.16-29, February, 1989.

[20] Hanqi Zhuang, Zvi S. Roth and R. Sudhakar. "Simultaneous robot/-world and tool/flange calibration by solving homogeneous transformation equations of the form AX = YB". *1994, Robotics and Automation, IEEE Transactions on*, vol.10, no.4, pp.549-554, August, 1994.

[21] Federico Vicentini, Nicola Pedrocchi, Matteo Malosio and Lorenzo Molinari Tosatti. "High-Accuracy Hand-Eye Calibration from Motion on Manifolds". *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3327-3334, 2011.

[22] Jan Heller, Didier Henrion and Tomas Pajdla. "Hand-eye and robot-world calibration by global polynomial optimization". *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pp.3157-3164, 2014.

[23] Jill Schmidt, Devin R. Berg and Heidi-Lynn Ploeg. "Precision, repeatability and accuracy of Optotrak® optical motion tracking systems". *Int. J. Experimental and Computational Biomechanics*, vol.1, no.1, pp.114-127, 2009.

[24] Paulo Debenest, Edwardo F. Fukushima, Yuki Tojo and Shigeo Hirose. "A New Approach to Humanitarian Demining. Part 2: Development and Analysis of Pantographic Manipulator". *Autonomous Robots*, vol.18, no.3, pp.323-336, 2005.

[25] Martin A. Fischler and Robert C. Bolles. "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography". *Communications of the ACM*, vol.24, Issue 6, pp.381-395, 1981.

[26] K. Levenberg. "A method for the solution of certain problems in least squares". *Quarterly of Applied Mathematics*, vol.2, pp.164-168, 1944.

[27] Donald W. Marquardt. "An algorithm for least-squares estimation of nonlinear parameters". *Journal of the Society for Industrial Application Mathematics*, vol.11, no.2, pp.431-441, 1963.

[28] J. Nocedal and S.J. Wright. "Numerical Optimization". *Springer*, New York, 1999.

[29] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "Speeded-Up Robust Features (SURF)", *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, 2008.

[30] N. Chernov and C. Lesort. "Least Squares Fitting of Circles". *Journal of Mathematical Imaging and Vision*, vol.23, Issue 3, pp.239-252, 2005.

[31] Xiaoyi Jiang and Dachuan Cheng. "Fitting of 3D circles and ellipses using a parameter decomposition approach". *3-D Digital Imaging and Modeling 2005*, pp.103-109, June. 2005.

[32] Mark A. Robertson, Sean Borman and Robert L. Stevenson, "Dynamic range improvement through multiple exposures", *In Proc. of the Int. Conf. on Image Processing (ICIP99)*, pp.159-163, 1999.

[33] Kaneko A. M., Marino M., Fukushima E. F.. "Humanitarian Demining Robot Gryphon: New Vision Techniques and Optimization Methods". *IEEE/RSJ Int. Conf. on Intelligent Robotics and Systems (IROS)*, pp.228-233, October, 2010.

[34] R. Correal, G. Pajares and J.J. Ruz. "Automatic expert system for 3D terrain reconstruction based on stereo vision and histogram matching". *Expert Systems with Applications*, vol. 41, Issue 4, pp.2043-2051, 2014.

[35] Paul E. Debevec and Jitendra Malik. "Recovering high dynamic range radiance maps from photographs". Proceedings of the 24th annual conference on Computer graphics and interactive techniques, pp.369-378, 1997.

[36] Mertens Tom, Kautz Jan and Van Reeth Frank. "Exposure Fusion". *Pacific Graphics*, pp.382-390, 2007.

[37] Mark A. Robertson and Sean Borman and Robert L. Stevenson. "Estimation-Theoretic Approach to Dynamic Range Enhancement Using Multiple Exposures". *Journal of Electronic Imaging*, vol.12, pp.219-228, 2003.

[38] R. Fattal, M. Agrawala, and S. Rusinkiewicz. "Multiscale shape and detail enhancement from multi-light image collections". *ACM Transactions on Graphics*, vol.26, no.3, Article 51, July, 2007.

[39] Gelfand Natasha, Adams Andrew, Park Sung Hee and Pulli Kari. "Multi-exposure Imaging on Mobile Devices". *Proceedings of the International Conference on Multimedia*, pp. 823-826, 2010.

[40] Barbara Zitova and Jan Flusser. "Image registration methods: a survey". *Image and Vision Computing*, vol.21, Issue 11, pp.977-1000, October, 2003.

[41] R. Szeliski. "Image alignment and stitching: A tutorial". *Foundations and Trends in Computer Graphics and Vision archive*, , vol.2, Issue 1, pp.1-104, January, 2006.

[42] Brown, M. and Lowe, D.G.. "Recognising panoramas". *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, vol.2, pp.1218-1225, October, 2003.

[43] M. Brown and D. G. Lowe. "Automatic Panoramic Image Stitching using Invariant Features". *International Journal of Computer Vision*, vol.74, no.1, pp.59-73, 2007.

[44] Michal Irani and P. Anandan. "Robust multi-sensor image alignment". *Computer Vision, 1998. Sixth International Conference on*, pp.959-966, January, 1998.

[45] Steve Mann and Rosalind W. Picard. "Video orbits of the projective group: a simple approach to featureless estimation of parameters". *Image Processing, IEEE Transactions on*, vol.6, no.9, pp.1281-1295, September, 1997.

[46] Bernardo E. Pires and Pedro M. Q. Aguiar. "Featureless global alignment of multiple images". *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol.1, pp.57-60, September, Genova, Italy, 2005.

[47] Pedro M. Q. Aguiar. "Unsupervised simultaneous registration and exposure correction". *Image Processing, 2006 IEEE International Conference on*, pp.361-364, October, 2006.

[48] Tomaszewska Anna and Mantiuk Radoslaw. "Image Registration for Multi-exposure High Dynamic Range Image Acquisition". *In: Proc. of International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pp.49-56, Plzen-Bory, Czech Republic, 2007.

[49] R.A.Lane and N.A.Thacker. "Tutorial: Overview of Stereo Matching Research". *Technical report, University of Manchester*, 1998.

[50] Daniel Scharstein and Richard Szeliski. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms". *International Journal of Computer Vision*, vol.47, no.1, pp.7-42, May, 2002.

[51] Ke Zhang, Jiangbo Lu and Gauthier Lafruit. "Cross-Based Local Stereo Matching Using Orthogonal Integral Images". *Circuits and Systems for Video Technology, IEEE Transactions on*, vol.19, no.7, pp.1073-1079, July, 2009.

[52] Asmaa Hosni, Michael Bleyer, Christoph Rhemann, Margrit Gelautz and Carsten Rother. "REal-time local stereo matching using guided image filtering". *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pp.1-6, July, 2011.

[53] Xing Mei, Xun Sun, Mingcai Zhou, shaohui Jiao, Haitao Wang and Xiaopeng Zhang. "On building an accurate stereo matching system on graphics hardware". *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pp.467-474, November, 2011.

[54] Yuri Boykov, Olga Veksler and Ramin Zabih. "Fast approximate energy minimization via graph cuts". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.23, no.11, pp.1222-1239, November, 2001.

[55] J. Marroquin, S. Mitter and T. Poggio. "Probabilistic Solution of Ill-Posed Problems in Computational Vision". *Journal of the American Statistical Association*, vol.82, no.397, pp.76-89, 1987.

[56] Ramin Zabih and John Woodfill. "Non-parametric Local Transforms for Computing Visual Correspondence". *In Proceedings of European Conference on Computer Vision*, pp.151-158, Stockholm, Sweden, May, 1994.

[57] Hirschmuller Heiko. "Stereo Processing by Semiglobal Matching and Mutual Information". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.30, no.2, pp.328-341, February, 2008.

[58] Istvan Haller and Sergiu Nedevschi. "Design of interpolation functions for subpixel-accuracy stereo-vision systems". *IEEE Trans. on Image Process.*, vol.21, no.2, pp.889-898, February, 2012.

Appendix A

Notation

$$R(x,\phi_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\phi_x) & -sin(\phi_x) \\ 0 & sin(\phi_x) & cos(\phi_x) \end{bmatrix}$$

$$R(y,\phi_y) = \begin{bmatrix} cos(\phi_y) & 0 & sin(\phi_y) \\ 0 & 1 & 0 \\ -sin(\phi_y) & 0 & cos(\phi_y) \end{bmatrix}$$

$$R(z,\phi_z) = \begin{bmatrix} cos(\phi_z) & -sin(\phi_z) & 0 \\ sin(\phi_z) & cos(\phi_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T(P_x, P_y, P_z) = [P_x \ P_y \ P_z]^T$$

Appendix B

Parameters calculation of Humanitarian Demining Robot system

Using the reference point positions measured by stereo vision camera in the motion sequence $m(m = 1, 2, 3)$, with the calibration method (Pure Rotation with Joint Constraint Method) described in this paper, in the camera coordinate frame, the estimated normal vector ${}^{C}\boldsymbol{n}_m^{SC}$ and center ${}^{C}\boldsymbol{C}_m^{SC}$ of the circular arc can be solved. Since the joint axis of the moved joint must pass through the center of the circular arc, from the mechanical structure, the theoretical rotation axis $\boldsymbol{n}_m^T$ and rotation center $\boldsymbol{C}_m^T$ of the motion sequence m are known. For the motion sequence 1, 2 and 3, they meet (B.1) to (B.6).

$$\boldsymbol{C}_1^T = R(x, (\theta_{Cx} + \pi/2))R(y, \theta_{Cy})R(z, (\theta_{Cz} + \pi)){}^{C}\boldsymbol{C}_1^{SC} + T(P_{Cx}, P_{Cy}, P_{Cz}) \tag{B.1}$$

$$\boldsymbol{n}_1^T = R(y, \theta_{Cy})R(z, (\theta_{Cz} + \pi)){}^{C}\boldsymbol{n}_1^{SC} \tag{B.2}$$

$$\boldsymbol{C}_2^T = R(x, (\theta_{Cx} + \pi/2))R(y, \theta_{Cy})R(z, (\theta_{Cz} + \pi)){}^{C}\boldsymbol{C}_2^{SC} + T(P_{Cx}, P_{Cy}, P_{Cz}) \tag{B.3}$$

$$\boldsymbol{n}_2^T = R(y, \theta_{Cy})R(z, (\theta_{Cz} + \pi)){}^{C}\boldsymbol{n}_2^{SC} \tag{B.4}$$

$$\boldsymbol{C}_3^T = R(x, \theta_2)(R(x, (\theta_{Cx} + \pi/2))R(y, \theta_{Cy})R(z, (\theta_{Cz} + \pi)){}^{C}\boldsymbol{C}_3^{SC} + T(P_{Cx}, P_{Cy}, P_{Cz})) \tag{B.5}$$

$$\boldsymbol{n}_3^T = R(x,\theta_2)R(x,(\theta_{Cx}+\pi/2))R(y,\theta_{Cy})R(z,(\theta_{Cz}+\pi))^C\boldsymbol{n}_3^{SC} \quad \text{(B.6)}$$

Note:

1. ${}^C\boldsymbol{C}_m^{SC} = [{}^CC_{xm}^{SC} \quad {}^CC_{ym}^{SC} \quad {}^CC_{zm}^{SC}]^T$, $m = 1,2,3$
2. ${}^C\boldsymbol{n}_m^{SC} = [{}^Cn_{xm}^{SC} \quad {}^Cn_{ym}^{SC} \quad {}^Cn_{zm}^{SC}]^T$, $|{}^C\boldsymbol{n}_m^{SC}| = 1$, $m = 1,2,3$
3. $\boldsymbol{C}_1^T = [\# \quad 0 \quad L_1]^T$, $\boldsymbol{n}_1^T = [-1 \quad 0 \quad 0]^T$
4. $\boldsymbol{C}_2^T = [\# \quad 0 \quad 0]^T$, $\boldsymbol{n}_2^T = [-1 \quad 0 \quad 0]^T$
5. $\boldsymbol{C}_3^T = [0 \quad 0 \quad \#]^T$, $\boldsymbol{n}_3^T = [0 \quad 0 \quad -1]^T$

After the normal vectors and centers of the circular arcs in the motion sequence 1, 2 and 3 are estimated, the calibration parameters $\boldsymbol{p} = \{\hat{\theta}_2, \hat{\theta}_3, \theta_{Cx}, \theta_{Cy}, \theta_{Cz}, P_{Cx}, P_{Cy}, P_{Cz}\}$ can be computed.

Firstly, with (B.2) and (B.4), the parameters θ_{Cy} and θ_{Cz} are solved.

Next, the parameters θ_{Cx},P_{Cx},P_{Cy},P_{Cz},$\hat{\theta}_2$ are calculated with (B.1), (B.3), (B.5) and (B.6).

Finally, with the internal reference point positions in motion sequence 1, by solving the optimization problem defined with (B.7), the parameter $\hat{\theta}_3$ is solved.

$$\min \sum_{t=1}^{N_1} |{}^R\boldsymbol{P}_{1,t}^{SC} - {}^R\boldsymbol{P}_{1,t}^{SE}|^2 \quad \text{(B.7)}$$

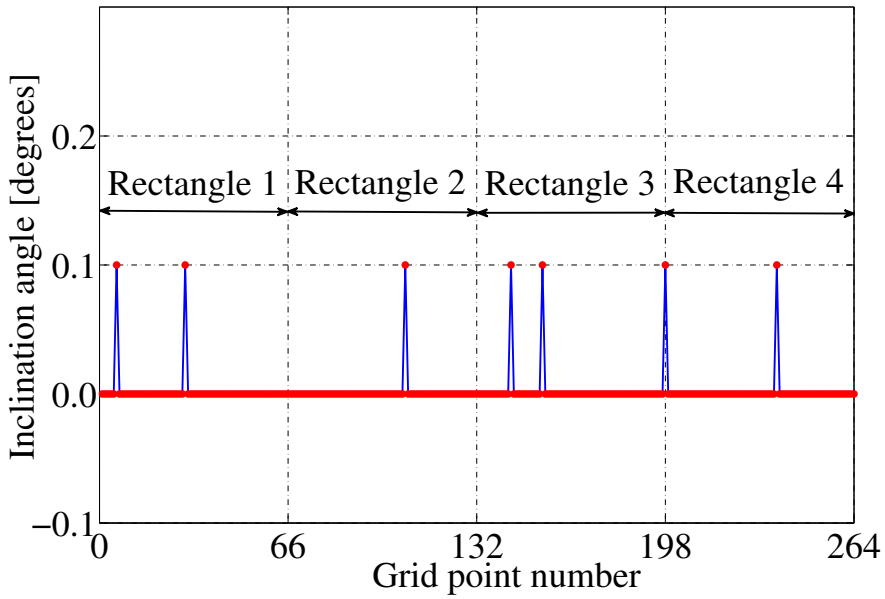
s.t.

$$\theta_3^L \le \hat{\theta}_3 \le \theta_3^H$$

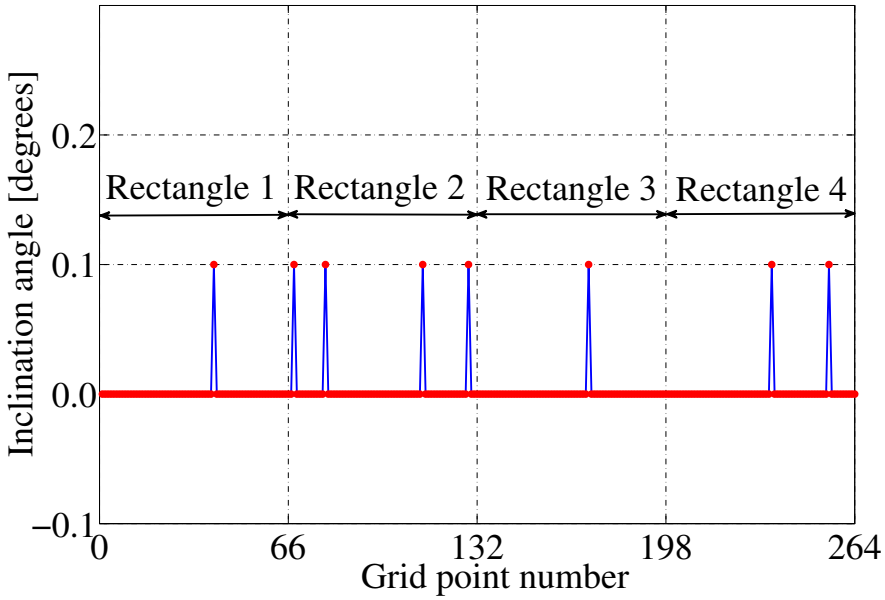
Appendix C

Manipulator inclination

In order to check the manipulator inclination, setting the joint initial angles to be the nominal values, for the marker on the manipulator's tip (Figure 2.8), in the robot coordinate frame, the trajectory of rectangle motion was generated with inverse kinematics. The rectangle motion with an area of 1000.0 mm x 2000.0 mm at different heights is shown in Figure 2.9. The digital angle gauge with the resolution of 0.1 degrees was placed on the manipulator base to measure the manipulator inclination angle in X direction and Y direction respectively. The results are shown in Figure C.1. From the experiment results, it is noticed that the inclination to the manipulator base was not larger than 0.1 degrees when moving the manipulator in its working space. When there is inclination of 0.1 degrees, the maximum position error at the manipulator's tip is 5.9 mm which meets the error requirement of Humanitarian Demining Robot system. So it is safe to not take into account the inclination effect.



(a) Manipulator inclination angles in X direction.



(b) Manipulator inclination angles in Y direction.

Figure C.1: Checking manipulator inclination.

Appendix D

Achievements

D.1 Papers in refereed journals

1. Jianhua Li, Gen Endo, Edwardo F. Fukushima. "Terrain Mapping under Extreme Light Conditions with Direct Stereo Matching Method through Aggregating Matching Costs by Weight", Journal of Advanced Robotics (under review).

2. Jianhua Li, Alex Kaneko, Gen Endo, Edwardo F. Fukushima. "In-Field Self-Calibration of Robotic Manipulator Using Stereo Camera - Application to Humanitarian Demining Robot -", Journal of Advanced Robotics (in press).

D.2 Papers in refereed conferences

1. Jianhua Li, Gen Endo, Edwardo F. Fukushima. "Hand-Eye Calibration using Stereo Camera through Pure Rotations -Fitting Circular Arc in 3D Space with Joint Angle Constraint-", 2015 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM2015), Busan, Korea, July, 2015.

2. Jianhua Li, Alex M. Kaneko, Edwardo F. Fukushima. "Proposal of Terrain Mapping under Extreme Light Conditions Using Direct Stereo Matching Methods", 2014 IEEE/SICE International Symposium on System Integration (SII2014), Tokyo, Japan, December, 2014.

D.3 Papers in non-refereed conferences

1. Jianhua Li, Edwardo F. Fukushima. "Proposal of a Wide View Image Sensor Configuration for Object Detection on the Water", 2P1-D01, JSME Robomech Conference 2015, Kyoto, Japan, May, 2015.

2. Jianhua Li, Alex M. Kaneko, Edwardo F. Fukushima. "Basic Indoor Experiment of 3D Mapping Using Stereo Vision Camera under Extreme Light Conditions", RSJ2014AC2A1-01. The 32nd annual conference of the Robotics Society of Japan (RSJ), Fukuoka, Japan, September, 2014.

3. Jianhua Li, Alex M. Kaneko, Edwardo F. Fukushima. "Humanitarian Demining Robot Gryphon Self-Calibration Using Stereo Vision Camera", 1A1-R11, JSME Robomech Conference 2013, Tsukuba, Japan, May, 2013.

D.4 Award

1. Best Paper Award. 2014 IEEE/SICE International Symposium on System Integration (SII2014), Tokyo, Japan, December, 2014.